



The attached material is posted on regulation2point0.org with permission.



J O I N T C E N T E R
AEI-BROOKINGS JOINT CENTER FOR REGULATORY STUDIES

QALYs and Policy Evaluation: A New Perspective

Matthew D. Adler^{*}

**Working Paper 05-01
January 2005**

This paper was published in the *Yale Journal of Health Policy, Law, and Ethics*.

^{*} The author is Matthew D. Adler, Professor of Law, University of Pennsylvania. The author thanks Craig Phillips for excellent research assistance, and Howard Chang, Chris Sanchirico, and Reed Shuldiner for very helpful comments and conversations. Please send comments to: madler@law.upenn.edu.



J O I N T C E N T E R

AEI-BROOKINGS JOINT CENTER FOR REGULATORY STUDIES

In order to promote public understanding of the impact of regulations on consumers, business, and government, the American Enterprise Institute and the Brookings Institution established the AEI-Brookings Joint Center for Regulatory Studies. The Joint Center's primary purpose is to hold lawmakers and regulators more accountable by providing thoughtful, objective analysis of relevant laws and regulations. Over the past three decades, AEI and Brookings have generated an impressive body of research on regulation. The Joint Center builds on this solid foundation, evaluating the economic impact of laws and regulations and offering constructive suggestions for reforms to enhance productivity and welfare. The views expressed in Joint Center publications are those of the authors and do not necessarily reflect the views of the Joint Center.

ROBERT W. HAHN

Executive Director

ROBERT E. LITAN

Director

COUNCIL OF ACADEMIC ADVISERS

KENNETH J. ARROW
Stanford University

MAUREEN L. CROPPER
University of Maryland

PHILIP K. HOWARD
Common Good

PAUL L. JOSKOW
Massachusetts Institute
of Technology

DONALD KENNEDY
Stanford University

ROGER G. NOLL
Stanford University

GILBERT S. OMENN
University of Michigan

PETER PASSELL
Milken Institute

RICHARD SCHMALENSEE
Massachusetts Institute
of Technology

ROBERT N. STAVINS
Harvard University

CASS R. SUNSTEIN
University of Chicago

W. KIP VISCUSI
Harvard University

All AEI-Brookings Joint Center publications can be found at www.aei-brookings.org

© 2005 by the author. All rights reserved.

Executive Summary

This Article presents a new, welfarist defense of the use of QALYs (quality adjusted life years) in policy evaluation. It challenges both the conventional wisdom among health economists that QALY-based analysis is dominated by traditional cost-benefit analysis (i.e., the sum of willingness-to-pay amounts) as well as the standard view of public health researchers that QALYs should function as the effectiveness metric in a cost-effectiveness analysis. Instead, the Article defends a nontraditional form of cost-benefit analysis, where QALYs are multiplied by a conversion factor, for example \$100,000 per QALY, and added to the monetized non-health effects of a policy. Part I of the Article surveys the current literature on QALYs. Parts II and III show that QALYs can be a more accurate measure of overall well-being than WTP amounts, under certain conditions, and argues that cognitive difficulties interfering with the measurement of WTP amounts can be circumvented by QALYs. Part IV describes the limitations of QALYs. Part V discusses the role that QALYs should play in welfarist policy analysis, given their strengths and limitations. In particular, it presents a pragmatic approach to determining the QALY-to-dollar conversion factor, and sheds new light on the controversy about pricing whole lives versus life-years.

QALYs and Policy Evaluation: A New Perspective

Matthew D. Adler, J.D.*

INTRODUCTION

The “quality-adjusted life year” (QALY) is a metric for health and longevity that is now widely used by health economists, public health scholars, and others researching the economics of health care.¹ QALYs work like this: Imagine a life history or “profile” of health states $h_1, h_2 \dots h_n$, where each state h_j persists for t_j years. A health state can be death, perfect health, or any disease condition in-between: angina, bronchitis, lung cancer, depression, headaches, heart disease, and so on. Patients who have experienced the states, physicians familiar with the states, or members of the general population will have been surveyed and asked to rank each state h_j on a 0-1 scale of health quality, with 1 representing perfect

* Professor of Law, University of Pennsylvania. Many thanks to Craig Phillips for excellent research assistance, and to John Broome, Howard Chang, Robert Hahn, James Hammitt, Eric Posner, Chris Sanchirico, Amy Sinden, Cass Sunstein, and Reed Shuldiner for very helpful comments and conversations. This Article was written prior to the publication of JOHN BROOME, *WEIGHING LIVES* (2004), a rich and systematic philosophical treatment of the problem of measuring well-being and the impact of death on well-being. I have therefore, unfortunately, been unable to incorporate Broome’s foundational discussion into my own, more applied, analysis.

1. Good overviews of the QALY approach to measuring health and longevity include OFFICE OF RES. & DEV., U.S. EPA, *HUMAN HEALTH METRICS FOR ENVIRONMENTAL DECISION SUPPORT TOOLS: LESSONS FROM HEALTH ECONOMICS AND DECISION ANALYSIS* (2001) [hereinafter U.S. EPA, *HUMAN HEALTH METRICS*]; Paul Dolan, *The Measurement of Health-Related Quality of Life for Use in Resource Allocation Decisions in Health Care*, in 1B *HANDBOOK OF HEALTH ECONOMICS* 1723 (A.J. Culyer & J.P. Newhouse eds., 2000); Robert Fabian, *The Qaly Approach*, in *VALUING HEALTH FOR POLICY* 118 (George Tolley et al. eds., 1994); James K. Hammitt, *QALYs Versus WTP*, 22 *RISK ANALYSIS* 985 (2002) [hereinafter Hammitt, *QALYs Versus WTP*]; Magnus Johannesson et al., *Outcome Measurement in Economic Evaluation*, 5 *HEALTH ECON.* 279 (1996); Robert M. Kaplan, *Utility Assessment for Estimating Quality-Adjusted Life Years*, in *VALUING HEALTH CARE: COSTS, BENEFITS, AND EFFECTIVENESS OF PHARMACEUTICALS AND OTHER MEDICAL TECHNOLOGIES* 31 (Frank A. Sloan ed., 1996); Graham Loomes & Lynda McKenzie, *The Use of QALYs in Health Care Decisionmaking*, 28 *SOC. SCI. & MED.* 299 (1989); and J. Brazier et al., *A Review of the Use of Health Status Measures in Economic Evaluation*, *HEALTH TECH. ASSESSMENT*, May 1999, at 1. A comprehensive guide is *QUALITY OF LIFE AND PHARMACOECONOMICS IN CLINICAL TRIALS* (Bert Spilker ed., 2d ed. 1996).

health and 0 representing death. There are various techniques for eliciting the quality ranking, $q(h_j)$, the two most popular being the “time-tradeoff” method and the “standard-gamble” method. The first method seeks to determine the respondent’s point of indifference between living y years with the condition h_j , and x years in perfect health (with x less than y), and assigns h_j the number x/y . The second method seeks to determine the respondent’s point of indifference between living a given period of time with the condition, and a gamble with probability p of living in perfect health for that same period of time and probability $1-p$ of dying instantaneously. Health state h_j is then assigned the indifference probability, p , as its quality ranking.

The QALY number for a health profile is calculated as the sum of the quality-weighted years spent in each of its component health states. For example, if some person’s life-history, absent a medical intervention, would consist of $h_1 \dots h_n$, and after intervention would improve to $h_1^* \dots h_n^*$, then the QALY measure of the first profile is $\sum_j q(h_j) \times t_j$, the QALY measure of the second is $\sum_j q(h_j^*) \times t_j^*$, and the QALY gain secured by the intervention is the difference between these two sums. Similarly, if the individual would live m years in perfect health absent intervention, and $m+x$ years in perfect health after intervention, the QALY gain is x QALYs. QALYs are evidently a powerful tool for measuring the impact of choices that affect morbidity, longevity, or both—not only the choices of physicians, hospitals, HMOs, and insurers, but also governmental choices, such as the FDA’s pharmaceutical licensure decisions, the regulation of toxins by the EPA or OSHA, or HHS’s choices about Medicare coverage.

To date, QALYs have been generally employed in cost-effectiveness studies.² In a cost-effectiveness study, both the health and non-health impacts of different health-affecting choices are determined. Non-health impacts are measured in dollars, but health impacts are measured using some nonmonetary scale (either a QALY scale, or some other scale, e.g., a disease-specific scale in the case where the health effects of the choices at issue are confined to a single disease). Cost-effectiveness ratios are then used to determine which choice should be undertaken. Alternatively, the choice which maximizes health given a fixed budget for non-health costs is selected.

Though QALYS were invented in the early 1970s,³ the use of QALYS—at least by researchers—has skyrocketed in recent years. Literally hundreds of health care cost-effectiveness studies now appear in academic journals every

2. See *infra* text accompanying notes 25-30 (discussing cost-effectiveness analysis and use of QALYs in that context).

3. See Kaplan, *supra* note 1, at 35 (crediting the invention of QALYs to a 1970 article by S. Fanshel and J.W. Bush); George W. Torrance & David Feeny, *Utilities and Quality-Adjusted Life Years*, 5 INT’L J. TECH. ASSESSMENT IN HEALTH CARE 559, 568 (1989) (same).

year.⁴ An increasing fraction of these employ QALYs. One recent review found that only one or two such studies were published annually during the 1980s and less than ten annually during the early 1990s, but that since 1997 roughly fifty “cost-utility” studies have appeared each year.⁵ This body of work has been accompanied by new surveys, including a massive survey of the general population performed a few years ago in England,⁶ and a burgeoning secondary literature on QALYs.⁷

To be sure, QALYs remain largely an academic tool, at least in the United States. Fifteen years ago, Oregon infamously relied upon, then abandoned, QALYs in deciding what treatments it would cover under Medicaid.⁸ Around the same time, HHS published, but never finalized, a proposal to use cost-effectiveness in Medicare coverage decisions.⁹ More generally, “cost-effectiveness analysis has been used by [U.S.] public sector agencies on a very limited basis” in the health policy area,¹⁰ by contrast with some foreign jurisdictions, such as Australia, Britain, Canada, and New Zealand, which now

4. See Anne Elixhauser et al., *Health Care CBA and CEA from 1991 to 1996: An Updated Bibliography*, 36 MEDICAL CARE MS1, MS6 (May Supp. 1998) (surveying health care literature and identifying 1792 published cost-effectiveness studies for the period 1991-96 and 1123 for the period 1979-90).

5. See Dan Greenberg & Joseph S. Pliskin, *Preference-Based Outcome Measures in Cost-Utility Analysis: A 20-Year Overview*, 18 INT’L J. TECH. ASSESSMENT 461, 463 (2002). Another study finds even greater numbers of published cost-utility studies. See Peter J. Neumann et al., *Growth and Quality of the Cost-Utility Literature, 1976-2001*, 8 VALUE IN HEALTH 3, 5 (2005).

6. See Paul Dolan et al., *The Time Trade-Off Method: Results from a General Population Study*, 5 HEALTH ECON. 141 (1996).

7. For bibliographies, see Dolan, *supra* note 1, at 1755-60; and U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 51. A useful list of references is provided at the end of each chapter in Brazier, *supra* note 1.

8. See, e.g., Kaplan, *supra* note 1, at 53-59 (discussing the Oregon episode).

9. See Medicare Program; Criteria and Procedures for Making Medical Services Coverage Decisions that Relate to Health Care Technology, 54 Fed. Reg. 4302, 4308-09 (proposed Jan. 30, 1989); Medicare Program; Procedures for Making National Coverage Decisions, 64 Fed. Reg. 22,619, 22,620 (Apr. 27, 1999) (withdrawing proposed rule); see also Medicare Program; Criteria for Making Coverage Decisions, 65 Fed. Reg. 31,124, 31,127 (May 16, 2000) (notice of intent to publish proposed rule establishing criteria for Medicare coverage decisions, which suggests QALYs as a possible measure of health benefits); Medicare Program; Revised Process for Making Medicare National Coverage Determinations, 68 Fed. Reg. 55,634, 55,634-35 (Sept. 26, 2003) (withdrawing intent to pursue rulemaking).

10. Frank A. Sloan & Henry G. Grabowski, *Introduction and Overview*, 45 SOC. SCI. & MED. 505, 508 (1997); see also Eric A. Posner, *Transfer Regulations and Cost-Effectiveness Analysis*, 53 DUKE L.J. 1067, 1105-06 (2003) (defending use of cost-effectiveness analysis to assess transfer regulations and stating that “[a]gencies have rarely performed cost-effectiveness analysis, and their occasional efforts have been inadequate”).

regularly evaluate the cost-effectiveness of pharmaceuticals or health technologies proposed for public reimbursement.¹¹ And private entities, such as hospitals, HMOs, or medical insurers, which in principle could incorporate QALYs into a wide range of their decisions, seldom do so at present.¹²

But there are signs of change. The FDA, over the last half-decade or so, has repeatedly relied on QALYs in its rulemakings, pioneering a new approach to QALY-based analysis that I will describe and defend below.¹³ The EPA, in its major arsenic and radon rulemakings during the late 1990s, specifically declined to employ a QALY-based monetary valuation of the morbidity- and mortality-reduction benefits of the rules.¹⁴ In other rulemakings around the same time, however, and more recently as well, the EPA has experimented with the so-called

11. See Nancy Devlin & David Parkin, *Does NICE Have a Cost-Effectiveness Threshold and What Other Factors Influence Its Decisions? A Binary Choice Analysis*, 13 HEALTH ECON. 437, 437 (2004); Paul P. Glasziou & Andrew S. Mitchell, *Use of Pharmacoeconomic Data by Regulatory Authorities*, in QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 114.

12. See Bryan R. Luce & Ruth E. Brown, *The Use of Technology Assessment by Hospitals, Health Maintenance Organizations, and Third-Party Payers in the United States*, 11 INT'L J. TECH. ASSESSMENT IN HEALTH CARE 79, 85 (1995); Peter Neumann, *Why Don't Americans Use Cost-Effectiveness Analysis?*, 10 AM. J. MANAGED CARE 308, 308 (2004); Elaine J. Power & John Eisenberg, *Are We Ready To Use Cost-Effectiveness Analysis in Health Care Decision-Making? A Health Services Research Challenge for Clinicians, Patients, Health Care Systems, and Public Policy*, 36 MED. CARE MS10, MS11-12 (May Supp. 1998). But see Bernard S. Bloom, *Use of Formal Benefit/Cost Evaluations in Health System Decision Making*, 10 AM. J. MANAGED CARE 329, 332-33 (2004) (reporting results of a 2002 survey of public and private health care organizations, and finding that a majority of the private respondents employ cost-benefit or cost-effectiveness analysis). For general discussions of the use of QALYs by governmental agencies or private actors, see COST-EFFECTIVENESS IN HEALTH AND MEDICINE 18-20 (Marthe Gold et al. eds., 1996); U.S. CONGRESS OFFICE OF TECH. ASSESSMENT, IDENTIFYING HEALTH TECHNOLOGIES THAT WORK 122-30 (1994); Bloom, *supra*; Glasziou & Mitchell, *supra* note 11; Luce & Brown, *supra*; Power & Eisenberg, *supra*; Sloan & Grabowski, *supra* note 10; Neumann et al., *supra* note 5; and Frank A. Sloan & Christopher J. Conover, *The Use of Cost-Effectiveness/Cost-Benefit Analysis in Actual Decision Making: Current Status and Prospects*, in VALUING HEALTH CARE, *supra* note 1, at 207. Two important works that were published subsequent to the drafting of this Article, and that discuss the use of QALYs or cost-effectiveness analysis, are: PETER J. NEUMANN, USING COST-EFFECTIVENESS ANALYSIS TO IMPROVE HEALTH CARE: OPPORTUNITIES AND BARRIERS (2005); and Symposium, *Cost-Effectiveness Analysis in U.S. Healthcare Decision-Making: Where Is It Going?*, 43 MED. CARE II-1 (July Supp. 2005).

13. See *infra* text accompanying notes 192-201.

14. See National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, 66 Fed. Reg. 6976, 7044 (Jan. 22, 2001); National Primary Drinking Water Regulations; Radon-222, 64 Fed. Reg. 59,246, 59,337 (Nov. 2, 1999).

“value of statistical life year” (VSLY) approach to monetizing mortality, an approach closely related to QALYs.¹⁵ The D.C. Circuit opinion in the *American Trucking* case, after striking down a provision of the Clean Air Act on nondelegation grounds, suggested that the EPA might cure those constitutional difficulties by measuring the benefits of air pollution regulations on a QALY scale.¹⁶ The HHS, in a recent rulemaking facilitating flu and pneumonia vaccinations, cited the cost-effectiveness of these vaccinations in promoting “year[s] of healthy life”—a synonym for a QALY.¹⁷ The Office of Management and Budget’s (OMB) current guide for Executive Order 12,866, which requires formal regulatory analyses of major rules, stipulates that these documents must include a cost-effectiveness analysis in the case of rules targeted at public health and safety,¹⁸ and it gives a qualified endorsement to QALYs as an appropriate effectiveness metric.¹⁹ The Public Health Service’s “Healthy People” initiative, an informational program which measures progress towards public health goals, employs QALYs as one of its key metrics.²⁰

In short, large amounts of money, time, and brainpower are currently being expended on QALY research, and that research now promises (or threatens, if you prefer) to shape policy. Legal scholars and policy analysts outside the health field should therefore pay more attention than they have done to QALYs. Are

15. See Control of Air Pollution from New Motor Vehicles: Heavy Duty Engine and Vehicle Standards and Highway Diesel Fuel Sulphur Control Requirements, 66 Fed. Reg. 5002, 5104 (Jan. 18, 2001); Control of Air Pollution from New Motor Vehicles: Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements, 65 Fed. Reg. 6698, 6784-87 (Feb. 10, 2000); Findings of Significant Contribution and Rulemaking on Section 126 Petitions for Purposes of Reducing Interstate Ozone Transport, 65 Fed. Reg. 2674, 2721-22 (Jan. 18, 2000); U.S. EPA, TECHNICAL ADDENDUM: METHODOLOGIES FOR THE BENEFIT ANALYSIS OF THE CLEAR SKIES INITIATIVE 35-37 (2002) [hereinafter U.S. EPA, TECHNICAL ADDENDUM]; Laura J. Lowenstein & Richard J. Revesz, *Anti-Regulation Under the Guise of Rational Regulation: The Bush Administration’s Approaches to Valuing Human Lives in Environmental Cost-Benefit Analyses*, 34 ENVTL. L. REP. 10,954 (2004) (describing the EPA’s use of VSLYs).

16. See *Am. Trucking Ass’n v. U.S. EPA*, 175 F.3d 1027, 1039-40 (D.C. Cir. 1999), *rev’d in part*, 531 U.S. 457 (2001) (finding that the Act was not an unconstitutional delegation).

17. Medicare and Medicaid Programs; Conditions of Participation: Immunization Standards for Hospitals, Long-Term Care Facilities, and Home Health Agencies, 67 Fed. Reg. 61,808, 61,813-14 (Oct. 2, 2002).

18. See OFFICE OF MGMT. & BUDGET, CIRCULAR A-4, at 9 (2003), *available at* <http://www.whitehouse.gov/omb/circulars/a004/a-4.pdf>.

19. See *id.* at 12-14.

20. See PENNIFER ERICKSON ET AL., U.S. DEP’T OF HEALTH & HUMAN SERVS., HEALTHY PEOPLE 2000 STATISTICAL NOTES NO. 7 (1995) (discussing “years of health life” measure), *available at* <http://www.cdc.gov/nchs/data/statnt/statnt/statnt07.pdf>; U.S. DEP’T OF HEALTH & HUMAN SERVS., HEALTHY PEOPLE 2010: UNDERSTANDING AND IMPROVING HEALTH 10 (2000) (same), *available at* <http://www.healthypeople.gov/document/pdf/uih/2010uih.pdf>.

QALYs more than a mere artifact of the resistance of physicians and public health officials to cost-benefit analysis? Is there a good reason, other than mere squeamishness about the monetary pricing of life and health that inheres in cost-benefit analysis, to incorporate QALYs in policy evaluation?

This Article addresses that question and answers it, affirmatively, in a novel way.²¹ My focus is governmental rather than private sector decision-making. I shall adopt a welfarist approach to policy analysis but—building on my prior work with Eric Posner²²—shall advance a conception of welfarism quite different from that held by many welfare economists, particularly the health economists who write about QALYs. Many welfare economists, and certainly many health economists, continue to structure their work around the criterion of Kaldor-Hicks efficiency²³—a criterion that, Posner and I have argued, should be rejected. Instead, welfarists should care about overall well-being. Overall well-being and Kaldor-Hicks efficiency are distinct criteria.

Traditional cost-benefit analysis, understood as the sum of willingness-to-pay/accept (WTP/WTa) amounts, is often seen by applied economists and policy analysts as a way to implement Kaldor-Hicks efficiency. Whatever the connection here,²⁴ traditional cost-benefit analysis is at best an imperfect proxy for overall well-being. Under certain conditions, QALY measures can better

21. Richard Pildes and Cass Sunstein have explored the role of QALYs in public policy, but their approach is different from mine. They see QALYs as an alternative to monetization, while I conceptualize QALYs as the input to a nontraditional, but monetized, cost-benefit analysis. See Richard H. Pildes & Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. CHI. L. REV. 1, 83-85 (1995).

22. See MATTHEW D. ADLER & ERIC A. POSNER, NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS (forthcoming 2006) [hereinafter ADLER & POSNER, COST-BENEFIT ANALYSIS]; Matthew D. Adler & Eric A. Posner, *Implementing Cost-Benefit Analysis when Preferences are Distorted*, in COST-BENEFIT ANALYSIS: LEGAL, ECONOMIC, AND PHILOSOPHICAL PERSPECTIVES 269 (Matthew D. Adler & Eric A. Posner eds., 2000) [hereinafter Adler & Posner, *Implementing Cost-Benefit Analysis*]; Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165 (1999) [hereinafter Adler & Posner, *Rethinking Cost-Benefit Analysis*].

23. A policy is Kaldor-Hicks efficient, relative to some alternative, if there is a hypothetical costless redistribution of wealth, from those whose welfare is increased by the policy to those whose welfare is reduced, which would make the policy Pareto-efficient, i.e., no one would be worse off and some would be better off.

24. Traditional cost-benefit analysis is not a perfect measure of Kaldor-Hicks efficiency given the Boadway paradox. See ROBIN BOADWAY & NEIL BRUCE, WELFARE ECONOMICS 263-71 (1984). Since I deny the moral significance of Kaldor-Hicks efficiency and see traditional cost-benefit analysis as a decision procedure that implements overall well-being rather than Kaldor-Hicks efficiency, it is not important for my purposes to determine whether the Boadway paradox, in practice, creates a large or only a small gap between cost-benefit analysis and Kaldor-Hicks efficiency.

track overall welfare than do WTP/WTa measures. In addition, the cognitive difficulties that ordinary individuals have in processing probabilities, and in trading off life and health for money, mean that the WTP/WTa amounts that individuals express in contingent-valuation studies or reveal through their behaviors may deviate substantially from the true WTP/WTa amounts—the amounts that genuinely track the individuals' welfare. QALY interviews and QALY aggregation can partly circumvent these cognitive failures.

So QALYs do have a place in welfarist policy evaluation, but not as the outcome measure in cost-effectiveness analysis. Rather, welfarist policy evaluation properly employs a nontraditional kind of cost-benefit analysis—one that monetizes certain impacts, such as health impacts, using valuation constructs other than WTP/WTa. Cost-benefit analysis, in the nontraditional form that I describe and defend in this Article, remains a technique that monetizes the welfare impact of governmental choices on affected individuals, then aggregates those monetary sums. But the use of QALY-to-dollar conversions, rather than WTP/WTa amounts derived from contingent valuation or revealed preference studies, is sometimes the best welfarist approach to monetizing welfare impacts.

Part I of the Article surveys existing scholarship about the role of QALYs in policy evaluation, split between extrawelfarists and welfarists. Extrawelfarists see QALYs as a measure of “healthiness” rather than well-being, while welfarists either support QALY-based cost-effectiveness analysis for the pragmatic reason that many in the health care field are averse to cost-benefit analysis, or reject QALYs altogether.

Part II and III present a novel defense of QALYs. Part II outlines the construct of overall well-being, understood (as per John Harsanyi) as the sum of utility numbers assigned by impartial spectators to individual life histories. I shall call these utility numbers “lifetime welfare units” (LWUs). Part III argues that QALYs are sometimes better proxies for LWUs than WTP/WTa amounts, for both conceptual and cognitive reasons.

But QALYs are hardly a perfect welfarist measuring rod. The LWU value of a life-history is a product of both its health and non-health attributes. QALYs capture only the first. Further, the overall health utility of a life-history may not be perfectly decomposable into the sum of the quality-adjusted durations of its component health states. These difficulties with QALYs are described in Part IV.

Part V describes the role that QALYs should play in cost-benefit analysis, given both the strengths and limitations of QALYs. It argues that QALY measures of health and longevity impacts should be converted into dollar amounts using a conversion factor, such as \$100,000 per QALY. Part V discusses how to set an optimal conversion factor. And it specifies the conditions under which QALY-to-dollar conversions should be used as inputs to cost-benefit analysis in lieu of WTP/WTa amounts.

Part V also provides a concrete illustration of my approach, on a topic of

much contemporary interest: the debate between those who argue that agencies should use a monetary value of life that is insensitive to (or at least not proportional to) the life expectancy of the persons whose lives are saved or lost, and those who contend that agencies should price life-years rather than lives. I argue for the pricing of life-years, at least under some circumstances. Many economists have criticized that approach, because it is at odds with traditional cost-benefit analysis. But once cost-benefit analysis is understood as a pragmatic technique that need not always rely on WTP/WTB measures, the life-years approach may be vindicated.

I. QALYS: THE CURRENT VIEW

This Part briefly reviews the current scholarly understanding of QALYs. Readers acquainted with this scholarship may want to proceed directly to Part II, but others are likely to find this review helpful as a point of departure. My own, novel approach to QALYs, presented in Parts II through V below, is best understood by contrast with the current literature.

Existing scholarship conceptualizes QALYs as an effectiveness metric for the purposes of a health policy cost-effectiveness analysis.²⁵ This is true both of scholars who adopt an “extrawelfarist” approach to health policy and of those who adopt a “welfarist” approach—a distinction to which I will return shortly. The literature on QALYs is characterized by various debates and divisions, but concurs in seeing QALY measures as a component of cost-effectiveness analysis.

Cost-effectiveness analysis of health policy is best described as a set of decision-analytic techniques that measure non-health effects in dollars but measure health and longevity on some nonmonetary scale, such as QALYs. Cost-effectiveness analysis has two main variants. The first variant assumes that non-health costs should be incurred up to some fixed budget, which is given exogenously, and picks the policy choice that maximizes health benefits within the budget. The second variant is not constrained by, nor intent on expending, an exogenous budget. Instead (roughly speaking) it excludes “dominated” choices (those that both produce smaller total health benefits than some alternative and are more expensive than that alternative); arrays the remaining choices in the order of their total health benefits, smallest to largest; and selects the choice furthest down the list whose incremental cost-effectiveness ratio does not exceed some cut-off ratio, for example \$50,000 per QALY.²⁶

25. On health care cost-effectiveness analysis generally, and cost-utility analysis specifically, see COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12; and MICHAEL F. DRUMMOND ET AL., METHODS FOR THE ECONOMIC EVALUATION OF HEALTH CARE PROGRAMMES 96-204 (2d ed. 1997).

26. This description is rough because the technique actually uses two different “dominance”

Appendix I provides an extended example of both variants of cost-effectiveness analysis. The first variant, maximizing health benefits for a fixed budget, is easy to understand. The second variant, which increases health production until the incremental cost-effectiveness ratio reaches some cut-off level, is a bit harder to grasp. One way to characterize this second variant of cost-effectiveness analysis is as follows: This technique reaches the same result that traditional cost-benefit analysis would *if* individual willingness-to-pay/accept (WTP/WTa) for a QALY were a constant value and the cut-off ratio were set to equal this constant.²⁷

In reality, however, both variants of cost-effectiveness analysis deviate from traditional cost-benefit analysis (the sum of individual WTP/WTa amounts). The first variant obviously does, because it takes for granted that the fixed budget should be expended on health. It selects the choice that maximizes health within the budget, over the choices of not expending all or any of the budget, even if the monetized benefits of the health-maximizing choice are less than the monetized costs.

As for the second variant, the premise that makes this coincide with traditional cost-benefit analysis, namely the existence of a constant WTP/WTa per QALY, is counterfactual. WTP for a given QALY gain is in fact heterogeneous.²⁸ Wealthier individuals will tend to pay more for a given QALY

notions to exclude choices: ordinary dominance, as described above, and “weak” or “extended” dominance. Further, it is structured to accommodate the possibility that health policies can be pursued simultaneously rather than being mutually exclusive. See Magnus Johannesson, *The Relationship Between Cost-Effectiveness Analysis and Cost-Benefit Analysis*, 41 SOC. SCI. & MED. 483, 483-84 (1995); Magnus Johannesson & Milton C. Weinstein, *On the Decision Rules of Cost-Effectiveness Analysis*, 12 J. HEALTH ECON. 459, 460-62 (1993); Göran Karlsson & Magnus Johannesson, *The Decision Rules of Cost-Effectiveness Analysis*, 9 PHARMACOECONOMICS 113 (1996); *infra* App. I.

The second variant of cost-effectiveness analysis, unlike the first, relies on a cut-off value. There are various methods, suggested in the literature, to generate this cutoff value. For example, the analyst might consult tables that show the cost-per-QALY of a range of existing programs, and might choose a cutoff value in the middle of that range. Or, recognizing that willingness-to-pay per QALY is heterogeneous, the analyst might try to determine what individuals are (in some sense) on average willing to pay. For a discussion, see Hans-Georg Eichler et al., *Use of Cost-Effectiveness Analysis in Health-Care Resource Allocation Decision-Making: How Are Cost-Effectiveness Thresholds Expected To Emerge?*, 7 VALUE IN HEALTH 518 (2004); and Milton C. Weinstein, *From Cost-Effectiveness Ratios to Resource Allocation: Where To Draw the Line?*, in VALUING HEALTH CARE, *supra* note 1, at 91-96.

27. See Mohan V. Bala et al., *Conditions for the Near Equivalence of Cost-Effectiveness and Cost-Benefit Analyses*, 5 VALUE IN HEALTH 338, 339-40 (2002); Johannesson, *supra* note 26, at 485-86.

28. See Paul Dolan & Richard Edlin, *Is It Really Possible To Build a Bridge Between Cost-Benefit Analysis and Cost-Effectiveness Analysis?*, 21 J. HEALTH ECON. 827, 837-38 (2002); Alan

gain.²⁹ If health does not change the welfare effect (“marginal utility”) of material consumption, healthier individuals will tend to pay less for a given QALY gain. Where health does change the marginal utility of consumption, healthier individuals may pay more for a given QALY gain.³⁰ Individuals for whom consumption is relatively more important, and health less important, will tend to pay less for a given QALY gain than their less materialistic counterparts. All these points are true both if WTP is keyed to preferences and even if WTP is keyed to a conception of well-being which is allowed to deviate from preferences. Since health and consumption do not make linear contributions to well-being at a single, constant rate—on any plausible account of well-being, preferentialist or not—WTP per QALY is heterogeneous. Thus the second variant of cost-effectiveness analysis can reach different results than traditional cost-benefit analysis. Appendix I may be helpful here, as it provides a concrete example that shows how both variants of cost-effectiveness analysis can diverge from traditional cost-benefit analysis.

Given the deviation between QALY-based policy evaluation, in the form of cost-effectiveness analysis, and traditional cost-benefit analysis, why use QALYs? The literature provides two answers to this question, “extrawelfarist” and “welfarist.”³¹ The extrawelfarist view sees health as a measurable characteristic of each person, distinct from that person’s well-being; conceptualizes QALYs as the measure of health; and argues that health programs should aim at some health-related goal (be it maximizing aggregate population QALYs, equalizing lifetime QALYs across the population, or something in between) rather than welfarist goals such as Kaldor-Hicks efficiency or overall well-being.

The best-known proponent of the extrawelfarist view is A.J. Culyer, who—drawing on scholarship by Amartya Sen—sees health as a “capability.”³² Sen’s

M. Garber & Charles E. Phelps, *Economic Foundations of Cost-Effectiveness Analysis*, 16 J. HEALTH ECON. 1, 28-29 (1997); Johannesson, *supra* note 26, at 486-87; *see also* Bala et al., *supra* note 27, at 344-45 (finding that constant WTP per QALY is sufficient but not necessary for the equivalence of cost-effectiveness and cost-benefit analysis, and identifying weaker necessary conditions that “are unlikely to hold in practice.”). For a recent empirical study confirming heterogeneity of WTP per QALY, *see* Duska M. Franic et al., *Quality-Adjusted Life Years Was a Poor Predictor of Women’s Willingness To Pay in Acute and Chronic Conditions: Results of a Survey*, 58 J. CLINICAL EPIDEMIOLOGY 291, 301 (2005).

29. *See, e.g.*, William N. Evans & W. Kip Viscusi, *Income Effects and the Value of Health*, 28 J. HUM. RESOURCES 497, 498-99, 516 (1993).

30. *See id.* at 499-500 (noting that ill health can increase or decrease the marginal utility of income).

31. *See* Dolan, *supra* note 1, at 1727-29.

32. *See, e.g.*, A.J. Culyer, *Commodities, Characteristics of Commodities, Characteristics of*

work, in turn, is part of a larger philosophical school, including such luminaries as John Rawls and Ronald Dworkin, which suggests that the “currency of justice” consists in something other than welfare.³³ Rawls famously argues that distributive justice concerns the distribution of “primary goods,” rather than well-being. Dworkin argues that genuine equality between persons involves their starting from a position of equal “resources,” not attaining equal welfare levels. Sen’s notion of a “capability” is a variation on Rawlsian primary goods and Dworkinian “resources.”

Extrawelfarists such as Culyer have a ready answer for why health policy analysts should continue to use cost-effectiveness analysis or some other form of QALY-based analysis rather than cost-benefit analysis: namely that health, rather than welfare, ought to be the underlying concern of health policy.³⁴ Cost-effectiveness analysis, as seen by the extrawelfarist, is a technique for maximizing the population’s health. The fact that the health-maximizing choice need not be the same as the welfare-maximizing choice or the Kaldor-Hicks efficient choice is no surprise, and no demerit for this technique, since health and welfare are different.

Is the extrawelfarist defense of QALY-based analysis persuasive? The extrawelfarist is plausibly correct that health is a measurable characteristic of individual lives which is distinct from welfare.³⁵ For example, restoring a paraplegic to full mobility would dramatically improve her health, but might not dramatically improve her welfare if she is happy as a paraplegic and her lifestyle with full mobility would remain sedentary.³⁶

Whether health rather than welfare should be the concern of health care policy and, more generally, whether the maximization or fair distribution of “capabilities” or “resources” or “primary goods” rather than welfare should be the focus of social planning and policy evaluation, are thorny questions that I will

People, Utilities, and the Quality of Life, in *QUALITY OF LIFE: PERSPECTIVES AND POLICIES* 9 (Sally Baldwin et al. eds., 1990) [hereinafter Culyer, *Commodities*]; A.J. Culyer, *The Normative Economics of Health Care Finance and Provision*, 5 *OXFORD REV. ECON. POL’Y* 34, 50-55 (1989). Sen has developed the capability view in many works, including AMARTYA SEN, *Equality of What?*, in *CHOICE, WELFARE AND MEASUREMENT* 353 (1982). For a recent symposium discussing the application of the capability approach to health policy, see Symposium, *Equity, Capabilities and Health*, 60 *SOC. SCI. & MED.* 219 (2005).

33. This extrawelfarist school is surveyed and criticized in Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 212, 215-16.

34. See, e.g., Bala et al., *supra* note 27, at 345; Dolan & Edlin, *supra* note 28, at 838.

35. See Culyer, *Commodities*, *supra* note 32, at 14-15 (distinguishing between arthritic’s health need and her marginal or total utility).

36. Cf. Philip Brickman et al., *Lottery Winners and Accident Victims: Is Happiness Relative?*, 36 *J. PERSONALITY & SOC. PSYCH.* 917, 920-21 (1978) (finding that paraplegic and quadriplegic accident victims were surprisingly happy).

not address in detail here. Richard Arneson and others have ably defended the view that welfare is the currency of justice,³⁷ and I have elsewhere made my own modest contribution to this deep normative debate.³⁸ Sometimes, extrawelfarism is linked to egalitarianism.³⁹ But that linkage is a mistake, since welfarist views can be as egalitarian as you like, and reciprocally extrawelfarist views need not be particularly egalitarian. More concretely, there is nothing egalitarian about cost-effectiveness analysis, which will choose a program that benefits healthy individuals over one that benefits sick individuals if further health improvements for the healthy are less expensive.⁴⁰

Sometimes, instead, a linkage is drawn between extrawelfarism and the morally attractive notions of responsibility and desert.⁴¹ Simple welfarism ignores the fact that a given individual might be responsible for a shortfall in her welfare. But welfarism need not be so simplistic.⁴² Conversely, traditional cost-effectiveness analysis is no more sensitive to responsibility and desert than traditional cost-benefit analysis. Cost-effectiveness analysis will channel social resources to individuals who are responsible for their poor health states or high mortality risks, and away from more deserving types, if the poor health or high risk of the less deserving is cheaper to remedy.

I will not belabor these arguments here, but will simply take as given the welfarist framework for policy evaluation. To be sure, even if welfarists are correct as a moral matter—even if it is true that governmental choice morally ought to track some function of welfare, rather than some function of health or other capabilities/resources/primary goods—it remains the case that welfarism is legally precluded in certain domains. For example, the Clean Air Act requires the EPA to set air quality standards that “are requisite to protect the public health”

37. See, e.g., Richard J. Arneson, *Welfare Should Be the Currency of Justice*, 30 CAN. J. PHIL. 497 (2000); Andrew Moore & Roger Crisp, *Welfarism in Moral Theory*, 74 AUSTRALASIAN J. PHIL. 598, 613 (1996) (defending welfarism against influential criticisms, and concluding that “welfarism [is] a powerful and attractive position”).

38. See Matthew D. Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 FL. ST. U. L. REV. 241, 307-09 (2000) [hereinafter Adler, *Beyond Efficiency and Procedure*]; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 215-16.

39. See Dolan, *supra* note 1, at 1727 (“[Extrawelfarists] will typically focus attention on equality of health . . .”).

40. See Peter A. Ubel et al., *Improving Value Measurement in Cost-Effectiveness Analysis*, 38 MED. CARE 892, 893-94 (2000) (noting that QALY-based cost-effectiveness analysis ignores the distribution of health benefits, and in particular has been criticized for not giving sufficient priority to interventions that are life-saving interventions, benefit those who are severely ill, or benefit those with limited treatment potential due to disability or chronic illness).

41. See Arneson, *supra* note 37, at 504-05.

42. See *id.* at 506-08.

and “allow[] an adequate margin of safety.”⁴³ The Occupational Safety and Health Act stipulates that OSHA shall regulate toxic substances in the workplace so as to ensure, to the extent feasible, that no worker suffers “material impairment of health or functional capacity.”⁴⁴ These provisions evidently give special priority to health. They would therefore seem to preclude cost-benefit analysis (both the traditional form and the nontraditional or “hybrid” variant this Article will defend), and indeed have been read to preclude cost-benefit analysis by the Supreme Court.⁴⁵ Yet, as Cass Sunstein has demonstrated, there are plenty of regulatory domains in which health, safety, and environmental agencies such as the EPA, National Highway Traffic Safety Administration (NHTSA), OSHA, and the FDA are statutorily permitted and perhaps even required to engage in cost-benefit analysis.⁴⁶ In these domains, the extrawelfarist has no particularly strong legal leg to stand on, nor (as I believe Arneson has shown) does she have a particularly strong moral leg either.

In short: Do QALYs have a role to play in *welfarist* policy evaluation? The remainder of this Article will focus on this question, and place the broader debate between welfarists and extrawelfarists to one side. The welfarists who have written about QALYs, mainly health economists, divide into two groups. One group recognizes that cost-effectiveness analysis can diverge from traditional cost-benefit analysis and therewith Kaldor-Hicks efficiency, but offers a pragmatic defense of QALYs: The health care field will not accept cost-benefit analysis. For example, Alan Garber explains:

[Cost-benefit] analysis requires placing dollar valuations on the outcomes of any program or intervention To many in the worlds of medicine and of public health, any attempt to place a value on a human life . . . is anathema. Thus most ‘economic’ evaluations in health care have applied [cost-effectiveness] analysis, which limits the analyst’s responsibility to providing information about the efficiency with which alternative strategies achieve health effects.⁴⁷

Similarly, *Cost-Effectiveness in Health and Medicine*, the influential report by an expert panel of economists and public health scholars commissioned by the Public Health Service, states:

Because of [cost-benefit analysis’s] explicit grounding in welfare-economic

43. 42 U.S.C. § 7409(b)(1) (2000).

44. 29 U.S.C. § 655(b)(5) (2000).

45. See *Whitman v. Am. Trucking Ass’n*, 531 U.S. 457, 464-71 (2001); *Am. Textile Mfg. Inst., Inc. v. Donovan*, 452 U.S. 490, 506-23 (1981).

46. See Cass R. Sunstein, *Cost-Benefit Default Principles*, 99 MICH. L. REV. 1651 (2001).

47. Alan M. Garber, *Advances in Cost-Effectiveness Analysis of Health Interventions* 4 (Nat’l Bureau of Econ. Research, Working Paper No. 7198, 1999).

principles, it is natural to ask why one would use cost-effectiveness rather than cost-benefit analysis if one wants to build from a welfare-economic foundation. Our interest in cost-effectiveness analysis derives largely from its broad acceptance within the health care field, in contrast to the skepticism that often greets cost-benefit analyses in that arena.

It is the distinguishing feature of [cost-benefit analysis] that offends some sensibilities: In [cost-benefit analysis], the benefit of the health intervention is expressed in dollar terms rather than in terms of a nonmonetary effectiveness measure.⁴⁸

Other health economists reject this pragmatic defense of QALYs and argue that, given the divergence between cost-effectiveness analysis and Kaldor-Hicks efficiency, QALYs should not be used in policy evaluation. Instead, they think, traditional cost-benefit analysis is the appropriate tool. For example, Mark Pauly explains how cost-benefit analysis implements Kaldor-Hicks efficiency (“the reason for preferring programs that maximize net benefits is [that] any such program can always be financed in such a way that everyone in society can be made better off”⁴⁹), describes efficiency as the foundation of welfare economics (“[the] potential compensation test . . . is the one welfare economics generally uses”⁵⁰) and concludes by rejecting cost-effectiveness analysis: “[In general] cost-effectiveness analysis is much less suitable . . . than cost-benefit analysis.”⁵¹ Similarly, Paul Dolan and Richard Edlin suggest that the only way to justify QALYs is by moving to extrawelfarism:

In showing that there is currently no meaningful link between [cost-benefit analysis] and [cost-effectiveness analysis], we have also shown that [cost-effectiveness analysis] is not currently justifiable on strictly welfarist grounds. Instead, [cost-effectiveness analysis] would seem to be justifiable only on non-welfarist grounds where the output of health care is judged according to its contribution to health itself, rather than according to the extent to which it contributes to overall welfare.⁵²

Or, to quote Don Kenkel: “[W]hen we accept the methodology of welfare economics, we should use cost-benefit analysis, not cost-effectiveness

48. COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12, at 28.

49. Mark V. Pauly, *Valuing Health Care Benefits in Money Terms*, in VALUING HEALTH CARE, *supra* note 1, at 101.

50. *Id.*

51. *Id.* at 111.

52. Dolan & Edlin, *supra* note 28, at 838.

analysis.”⁵³

Both of these positions are problematic. Consider first the welfarist defense of QALY-based analysis as a second-best decision procedure justified by the resistance of the relevant community (doctors and public health officials) to the first-best procedure, cost-benefit analysis. At least at the regulatory level, any norm against pricing life that might once have existed has long been dissipated. Federal agencies have published numerous cost-benefit analyses incorporating an explicit, monetary valuation of human life.⁵⁴ More than thirty-five years ago, welfare economists demonstrated how the WTP/WTa methodology could be employed to monetize fatalities, by asking for WTP/WTa for a change in the risk of death rather than for certain death or the avoidance of certain death.⁵⁵ This theoretical scholarship generated a vast body of empirical work, typically employing wage-risk studies to estimate a “value of statistical life”;⁵⁶ and these empirical and theoretical literatures, in turn, have now decisively influenced agency practice.

Pragmatic constraints are surely relevant to policy evaluation. If a strong taboo on pricing life were in place, then that might justify governmental agencies in employing a non-cost-benefit procedure that, albeit second-best, did not offend the taboo. But there is no such taboo, now, at least in the governmental context. And the anti-pricing norms that might still obtain in other contexts (for example, among physicians or hospital administrators) would presumably weaken or dissolve once cost-benefit analysis started to be practiced there.

Consider next the welfarist position that rejects QALYs because cost-benefit analysis, not cost-effectiveness analysis, tracks Kaldor-Hicks efficiency. This position is doubly problematic. First, Kaldor-Hicks efficiency itself lacks moral significance. To recapitulate very quickly arguments that Posner and I have set

53. Don Kenkel, *On Valuing Morbidity, Cost-Effectiveness Analysis, and Being Rude*, 16 J. HEALTH ECON. 749, 755 (1997).

54. See, e.g., Matthew D. Adler, *Fear Assessment: Cost-Benefit Analysis and the Pricing of Fear and Anxiety*, 79 CHI.-KENT L. REV. 977, 981 n.21 (2004) [hereinafter Adler, *Fear Assessment*] (finding that more than twenty-five percent of the cost-benefit analyses in the American Enterprise Institute database of major rulemakings for 1996-99 monetized death, illness, or injury).

55. See Michael W. Jones-Lee, *The Value of Changes in the Probability of Death or Injury*, 82 J. POL. ECON. 835 (1974); E.J. Mishan, *Evaluation of Life and Limb: A Theoretical Approach*, 79 J. POL. ECON. 687 (1971); T.C. Schelling, *The Life You Save May Be Your Own*, in PROBLEMS IN PUBLIC EXPENDITURE ANALYSIS 127 (Samuel B. Chase, Jr. ed., 1968).

56. For overviews of the theoretical and empirical scholarship on valuing life, see, for example, A. MYRICK FREEMAN III, *THE MEASUREMENT OF ENVIRONMENTAL AND RESOURCE VALUES: THEORY AND METHODS* 297-351 (2d ed. 2003); W. KIP VISCUSI, *FATAL TRADEOFFS: PUBLIC AND PRIVATE RESPONSIBILITIES FOR RISK* 17-74 (1992); W. Kip Viscusi & Joseph E. Aldy, *The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World*, 27 J. RISK & UNCERTAINTY 5 (2003).

forth in great detail elsewhere,⁵⁷ and that many others have made as well.⁵⁸ A Kaldor-Hicks efficient policy is merely one that could, hypothetically, be transformed into a Pareto-superior policy, through a costless redistribution from those who gain to those who lose. If the redistribution actually occurs, then the policy is actually (not just potentially) Pareto-superior and the criterion of Kaldor-Hicks efficiency is not needed to explain why the policy is attractive. If the redistribution doesn't occur, then the choice actually produces both winners and losers, and the link to Pareto-superiority has disappeared. Overall welfare, by contrast with Kaldor-Hicks efficiency, is a morally important notion; but a welfare-maximizing policy need not be potentially Pareto-superior, or vice versa. Because Kaldor-Hicks efficiency is no gold standard, the divergence between QALY-based analysis and the Kaldor-Hicks criterion is not an indictment of QALYs.

The welfarist position that rejects QALYs is problematic, too, because it hews to the widely shared but mistaken assumption that QALY-based analysis is necessarily a species of cost-effectiveness analysis. Once we shift from efficiency to the genuine gold standard—overall welfare—it will emerge that QALYs might function not as the measure of health in a cost-effectiveness analysis, but in a new analytic role.

I will ultimately argue below, in Part V, for a nontraditional, “hybrid” form of cost-benefit analysis in which health impacts are monetized by measuring those impacts on a QALY scale and then converting them to dollars using a QALY-to-dollar conversion factor.⁵⁹ Such QALY-to-dollar conversions should be used as a supplement to, and in some cases in lieu of, WTP/WTa measures of health impacts. The preferred tool for welfarist health policy evaluation should be neither traditional cost-benefit analysis alone nor cost-effectiveness analysis, but rather should include the “hybrid” cost-benefit technique just described.

This concrete recommendation will be my practical advice to policymakers. But it will not make much sense until we reconceptualize welfarist policy analysis in substantial, novel ways: by seeing overall welfare, not Kaldor-Hicks efficiency, as the underlying goal; and by understanding both QALYs and WTP/WTa amounts as workable but imperfect tools for measuring changes to overall welfare. Let us turn now to that reconceptualization.

57. See Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 248-59; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 22, at 273-74; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 187-94.

58. See Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 249 (citing critics of Kaldor-Hicks efficiency).

59. See *infra* Part V.

II. OVERALL WELL-BEING AND LWUS

This Part provides the philosophical foundations for my revisionary account of QALYs. It clarifies the construct of overall well-being, drawing upon John Harsanyi's notion of "extended" lotteries, specifically "extended" lotteries over life-histories. Overall well-being, thus construed, should be representable as the sum of "lifetime welfare units" (LWUs). LWUs are the true metric for interpersonal welfare—the true welfarist gold standard for evaluating governmental decisional techniques.

The discussion here is quite theoretical. But it is, I believe, absolutely essential to a proper understanding of QALYs. In order to see how QALYs and WTP/WTAs amounts are both imperfect proxies for LWUs—the thrust of Parts III and IV below—it is crucial to grasp the concept of an LWU.

Overall well-being is morally relevant, if not morally decisive.⁶⁰ Utilitarianism, the moral view that gives decisive weight to overall welfare, is often criticized because it ignores rights, distributive considerations, and "perfectionist" values such as the alleged value of preserving the environment quite apart from its welfare value. But none of these criticisms speaks to the weaker, pluralistic view that overall well-being is one criterion among others that together determine what ought morally to be done. Surely it is morally incorrect for government to reject a policy that increases overall welfare if the policy does not infringe any rights, distributive claims, or "perfectionist" values.⁶¹

There *is* one objection that applies to this weaker, pluralistic view as well as its more austere cousin, utilitarianism—namely, that interpersonal welfare comparisons are impossible and that talk about "overall well-being" is therefore meaningless. But this objection is misconceived. The scholarly skepticism about interpersonal comparisons that was widespread a half-century ago, tied to a broader skepticism about moral truth, has dissipated—certainly among philosophers and, to a substantial extent, among theoretical welfare economists too.⁶² Moral truth does not entail the existence of unnatural moral objects and properties.⁶³ Rather, moral assertions can plausibly be unpacked as assertions

60. This Part draws heavily on my prior work, particularly Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, and Adler, *Beyond Efficiency and Procedure*, *supra* note 38. I therefore generally cite to those works, rather than to the underlying scholarship in philosophy and economics, which they reference. My forthcoming book with Eric Posner, ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, will present a yet fuller account of the foundations of welfarist policy evaluation. The arguments in this Part are fleshed out in great detail in chapter 2 of the book.

61. See Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 302-19; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 209-16, 243-45.

62. See Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 204-09.

63. For an overview of "cognitivist" accounts of morality (accounts that recognize the existence of moral truths, moral beliefs, and moral facts) and their criticisms of noncognitivism, see

about the judgments or reactions of impartial and well-informed observers, and are true just in case such observers would indeed judge or react in the manner asserted.⁶⁴ Folk moral discourse presupposes the existence of moral truths, not just coincidentally, but because it ends up being extremely difficult to maintain a thoroughgoing moral skepticism. Consider the welfare economist who purports to be a moral skeptic but then claims that government “should” implement Pareto-superior policies. What is the status of that “should”?

Once moral skepticism goes by the wayside, skepticism about interpersonal comparisons also naturally disappears. Why would it be the case that murdering, raping or torturing me truly violates my moral rights, but there can never be any truth of the matter about whether a welfare gain for me exceeds a welfare loss for you? Note also that if interpersonal welfare comparisons are impossible, then a wide range of moral theories must be rejected out of hand—not just utilitarianism, and the softer pluralistic view I am defending here, but also the egalitarian view currently popular among philosophers known as “prioritarianism,”⁶⁵ as well as any other partly or wholly egalitarian moral view that requires a comparison of the welfare levels of different persons.⁶⁶ Welfarism, in all its plausible variants, just goes out the window.

How then should interpersonal welfare comparisons be understood? What does it mean to say that overall welfare in outcome *O* is truly greater than overall welfare in outcome *O**? One plausible approach derives from Adam Smith and has been endorsed in modern times by such scholarly luminaries as Kenneth Arrow, R.M. Hare, and Donald Davidson and formalized by John Harsanyi.⁶⁷ Harsanyi suggests that the two outcomes might be viewed as “extended”

generally ALEXANDER MILLER, AN INTRODUCTION TO CONTEMPORARY METAETHICS (2003).

64. *See id.* at 195-201.

65. Prioritarians claim that distributive and utilitarian considerations are fused in a single moral criterion, namely overall weighted welfare. Greater weight is given to welfare changes affecting individuals whose welfare levels are low, and less weight to changes affecting those whose welfare levels are high. *See, e.g.,* SHELLY KAGAN, NORMATIVE ETHICS 52-54 (1998). I have previously argued against prioritarianism, in favor of a moral view that incorporates ordinary, unweighted, overall welfare as one among a plurality of moral criteria, and Posner and I do so at greater length in our forthcoming book. ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, ch. 2; Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 309-11.

66. *See* ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, ch. 2.

67. *See* Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 206-08; Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 292-302; John C. Harsanyi, *Morality and the Theory of Rational Behavior*, in UTILITARIANISM AND BEYOND 39 (Amartya Sen & Bernard Williams eds., 1982); John A. Weymark, *A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism*, in INTERPERSONAL COMPARISONS OF WELL-BEING 255 (Jon Elster & John E. Roemer eds., 1991).

lotteries. Assume, for simplicity, that the same N individuals exist in each outcome. Then the extended lottery $(O; p_1, p_2 \dots p_N)$ offers a p_1 chance of assuming the identity of individual 1 in outcome O , a p_2 chance of assuming the identity of individual 2 in outcome O , and so on. The extended lottery $(O^*; q_1, q_2 \dots q_N)$ offers a q_1 chance of being individual 1 in outcome O^* , a q_2 chance of being individual 2 in outcome O^* , and so on. Harsanyi argues that individuals can have “extended” preferences—preferences over extended lotteries—and that humans who were fully informed, fully rational, and otherwise idealized (let us call these idealized humans “spectators” or “observers”) would have the same extended preferences. Harsanyi proposes to analyze interpersonal welfare comparisons as involving an equiprobability extended lottery—one where the probabilities of assuming the identities of different persons are equal. Intuitively, overall welfare embeds a certain kind of equality among persons: Your welfare has no greater weight than mine, just because it is yours, in determining overall welfare.⁶⁸ In Harsanyi’s schema, this deep impartiality is realized by stipulating the following: O is better for overall welfare than O^* just in case all spectators would prefer the extended lottery $(O; 1/N, 1/N \dots)$ to $(O^*; 1/N, 1/N \dots)$. There are N persons in each outcome and the chance of being any one is $1/N$.

Harsanyi’s schema is an idealization. Asking actual humans, with all their various cognitive limitations, to rank life-histories, let alone lotteries over life-histories, might be thought a ludicrous undertaking. In fact the undertaking may not be so ludicrous: There is a small but growing survey literature, an offshoot from the main body of QALY surveys, which seeks to elicit individuals’ preferences for temporally extended health profiles consisting of a sequence of different health states.⁶⁹ In any event, the Harsanyi schema is not offered here as a decision-procedure that governmental decision-makers seeing to maximize overall welfare should actually employ. Rather, its function is analytic: It tells us what overall well-being *means*. Harsanyi invokes the preferences of idealized spectators contemplating extended lotteries in an attempt to *define* overall

68. See KAGAN, *supra* note 65, at 49.

69. See Paul F.M. Krabbe & Gouke J. Bonsel, *Sequence Effects, Health Profiles, and the QALY Model*, 18 MED. DECISION MAKING 178 (1998); Miriam Kuppermann et al., *Can Preference Scores for Discrete States Be Used To Derive Preference Scores for an Entire Path of Events?*, 17 MED. DECISION MAKING 42 (1997); Jeffrey Richardson et al., *The Measurement of Utility in Multiphase Health States*, 12 INT’L J. TECH. ASSESSMENT IN HEALTH CARE 151 (1996); Anne Spencer, *A Test of the QALY Model When Health Varies over Time*, 57 SOC. SCI. & MED. 1697 (2003); Jonathan R. Treadwell, *Tests of Preferential Independence in the QALY Model*, 18 MED. DECISION MAKING 418 (1998). Indeed, the so-called HYE (“healthy years equivalent”) methodology for health policy analysis, a competitor to QALYs that some scholars have vigorously defended, requires individual valuations of temporally extended profiles. See Dolan, *supra* note 1, at 1729 (discussing HYE); Johannesson, *supra* note 1, at 286-88 (same).

welfare.⁷⁰ This should not seem so strange to anyone familiar with contemporary welfare economics or contemporary philosophical scholarship about welfare; appeals to “fully-informed” or otherwise idealized preferences are very common in these literatures.⁷¹

But how, then, *should* actual policymakers go about their business? The quick answer is that maximizing overall welfare in the Harsanyi sense poses the predicament of choice under uncertainty (uncertainty about which extended lotteries the spectators would prefer) by decision-makers (the actual humans who make policy choices) with various cognitive limits. The technique of cost-benefit analysis is one response to this predicament. It may not be the ideal technique, but it is (or seems to be) the best available technique given the state of policy science at the dawn of the twenty-first century.⁷² The trick is to design cost-benefit analysis so as to be sensitive to the Harsanyi construct and, more concretely, so as to incorporate whatever evidence we might have about what the preferences of fully-informed humans would be, without exceeding the cognitive capacities of actual policymakers or system-designers. So the field of policy analysis faces a difficult, but not insoluble, problem of specifying the optimal, feasible variant of cost-benefit analysis. Part V of the Article addresses this problem at somewhat greater length. The point I wish to emphasize here is that Harsanyi’s schema defines a moral notion, overall well-being, as an esoteric standard (one framed in terms of spectators’ extended preferences) which will never be transparent to us, but which we can do our best to grasp and apply in the design of actual governmental choice-procedures.

A more troubling feature of Harsanyi’s schema is his claim that spectators would have the same extended preferences. What would justify this claim? Our ordinary preferences might diverge, even with full information. I might prefer to eat chocolate ice cream, while you might prefer to eat vanilla ice cream. If so, it is hard to see why idealized extended preferences must converge: I might prefer, *ceteris paribus*, a life history in which the subject eats chocolate to one in which the subject eats vanilla, while you might have the opposite preference.⁷³

This flaw in Harsanyi’s account can be remedied, I have argued elsewhere, by appealing to *convergent* extended preferences.⁷⁴ *O* is better for overall welfare than *O** just in case the spectators would *all* prefer a lottery that delivers outcome *O* and a $1/N$ chance of being each individual in *O*, over a lottery that delivers

70. See Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 216-25 (distinguishing between moral criteria and morally justified decision procedures).

71. See *id.* at 203 & n.100.

72. See *id.* at 225-43.

73. See Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 293.

74. See *id.* at 297-300.

outcome O^* and a $1/N$ chance of being each individual in O^* . On this modification of Harsanyi's schema, the extent to which extended preferences converge under full information is an empirical issue, dependent on the facts about human nature; full convergence is not presupposed a priori.⁷⁵

In their seminal work on decision theory, von Neumann and Morgenstern proposed axioms for the rationality of preferences over lotteries which, if satisfied, will permit those preferences to be represented by "utility" numbers.⁷⁶ Given those axioms, an individual will prefer a lottery to another just in case it has greater utility, in turn calculated as the expected utility of its component parts.⁷⁷ In the Harsanyi setup, the component parts of the lotteries are what I have referred to as life-histories: having the identity of some individual i in outcome O . An equiprobability extended lottery can be represented as a lottery over life histories $\{L_1, L_2 \dots L_N\}$, each of which has probability $1/N$. The axioms of expected utility theory are, famously, violated in actual practice by untrained humans with their various biases and limitations.⁷⁸ But a strong case can be made that the axioms, if not descriptive, are normative—they are partly constitutive of rational choice—and thus would be satisfied by idealized spectators. So each spectator's extended preferences can be represented by a utility function W , such that the spectator prefers an equiprobability lottery over life-histories $\{L_1 \dots L_N\}$ to an equiprobability lottery over life-histories $\{L_1^* \dots L_N^*\}$ just in case: $1/N W(L_1) + 1/N W(L_2) + \dots 1/N W(L_N) > 1/N W(L_1^*) + 1/N W(L_2^*) + \dots 1/N W(L_N^*)$.⁷⁹ Since the same number of individuals exist in each outcome, and since each life history is given an equal probability, the $1/N$ factors drop out and it emerges that a given spectator prefers one extended lottery to a second just in case the sum of utility numbers for the life histories involved is greater in the first case: $W(L_1) + W(L_2) + \dots W(L_N) > W(L_1^*) + W(L_2^*) + \dots W(L_N^*)$.

Because (*pace* Harsanyi) extended preferences might not converge, the utility numbers tracking the various spectators' extended preferences might not be identical either. Formally, there is a family of interpersonal utility functions $W_1 \dots W_M$, corresponding to the M spectators; and one outcome O has greater

75. Indeed, respondents to QALY surveys do not fully converge on the weights they assign to health states. See U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 23; *infra* text accompanying notes 136-137.

76. See JOHN VON NEUMANN & OSKAR MORGENSTERN, THE THEORY OF GAMES AND ECONOMIC BEHAVIOR 15-31, 617-32 (3d ed. 1953).

77. On expected utility theory, see, e.g., SIMON FRENCH, DECISION THEORY: AN INTRODUCTION TO THE MATHEMATICS OF RATIONALITY 149-209 (1986); DAVID M. KREPS, NOTES ON THE THEORY OF CHOICE (1988); and MICHAEL D. RESNIK, CHOICES: AN INTRODUCTION TO DECISION THEORY 81-120 (1987).

78. See, e.g., SCOTT PLOUS, THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 84-188 (1993).

79. See Weymark, *supra* note 67, at 289-97.

overall welfare than another O^* just in case, for each of the utility functions $W_1 \dots W_M$, the sum of the utility numbers assigned to the component life-histories of O is greater than the sum of the utility numbers assigned to the component life-histories of O^* .⁸⁰ The technical term for this is a “supervaluation”: Overall well-being is a supervaluation over the extended utility functions of fully-informed spectators, conceiving each outcome as a lottery over life-histories with an equal chance of living each one. For simplicity, in the analysis of QALYs, WTPs, and cost-benefit analysis below, I will omit the supervaluationist caveat, assume convergence of extended preferences (as does Harsanyi), and refer to a single extended utility function W . But the argument carries over, I conjecture, to the more plausible case of partly divergent extended preferences. Because the analysis will prove to be long and complicated even on the assumption that spectators’ extended preferences fully converge, I will not attempt to verify the conjecture in this Article.

To sum up: Overall well-being has at least *prima facie* moral relevance. Talk of “overall well-being” presupposes the possibility of interpersonal welfare comparisons—but comparisons of welfare levels or differences are presupposed by a wide variety of moral views, skepticism about them is no more plausible than general moral skepticism, and Harsanyi’s construct of extended preferences and utility functions is a plausible way to give content to the notion of overall welfare. In the simple, limiting case, the spectators—our idealized counterparts—have convergent extended preferences, representable by a single utility function W . W assigns numbers to life-histories, and the overall well-being in some outcome is the sum of the utility numbers assigned to its component life-histories. The W function constitutes a cardinal and interpersonally comparable scale of welfare: a numerical representation of the degree to which changes in individual lives change the overall level of well-being.

Let us call the W -numbers “lifetime welfare units” (LWUs). I use this term, rather than “utility,” because “utility” is a very general concept that subsumes a host of numerical representations of preferences, including but not limited to the extended preferences of our idealized counterparts. Further, the word “lifetime” underscores that overall welfare is first and foremost a composite of whole lives, not momentary time-slices. Individual moments or periods within an individual life contribute to overall well-being by contributing to the goodness of the individual life-history of which they are a part. To assume otherwise—to stipulate that overall well-being is necessarily decomposable into the sum of momentary or periodic well-being—is to rule out the very possibility of

80. Cf. AMARTYA SEN, *INEQUALITY REEXAMINED* 46-49 (1992) (discussing an approach to social choice that focuses on convergent judgments and tolerates incompleteness).

sequencing effects.⁸¹ That would be deeply counterintuitive. Certain constituents of well-being, such as pains and pleasures, may not be subject to significant sequencing effects;⁸² but to insist a priori that this is true of all constituents seems wrongheaded.⁸³ As David Velleman explains:

Consider two different lives that you might live. One life begins in the depths but takes an upward trend Another life begins at the heights but slides downhill Surely, we can imagine two such lives as containing equal sums of momentary well-being. Your retirement is as blessed in one life as your childhood in the other; your nonage is as blighted in one life as your dotage in the other.

Yet even if we were to map each moment in one life onto a moment of equal well-being in the other, we would not have shown these lives to be equally good. For after the tally of good times and bad times had been rung up, the fact would remain that one life gets progressively better while the other gets progressively worse To most people, I think, the former story would seem like a better life story . . . in the sense that it is the story of a better life.⁸⁴

The notion of preferences over whole lives, rather than moments or periods, might seem odd. Yet this notion is accepted by the welfare-economic literature on inter-temporal choice, where the decomposability of utility functions over lives into the sum of utility functions over periods is understood as a contingent rather than necessary fact.⁸⁵ The same holds true of the QALY literature, where once more the additive formula $\sum_j q(h_{i,j}) \times t_j$ (the sum total of time alive, with each time span weighted for the quality of health in that period) is understood to be an accurate representation of a lifetime health history only if certain axioms are satisfied.⁸⁶

81. See *infra* text accompanying notes 160-163 (discussing empirical studies of sequencing effects).

82. See Daniel Kahneman, *Experienced Utility and Objective Happiness: A Moment-Based Approach*, in 1 THE PSYCHOLOGY OF ECONOMIC DECISIONS: RATIONALITY AND WELL-BEING 187, 191-94 (Isabelle Brocas & Juan D. Carrillo eds., 2003) (presenting an “objective happiness” account of the experiential component of well-being in which total experiential utility is a function of momentary experiential utility without sequencing effects).

83. See *id.* at 205 (“Objective happiness is not proposed as a comprehensive concept of human well-being, only as a significant constituent of it.”).

84. J. David Velleman, *Well-Being and Time*, in THE METAPHYSICS OF DEATH 329, 331 (John Martin Fischer ed., 1993).

85. See, e.g., Han Bleichrodt & Amiram Gafni, *Time Preference, the Discounted Utility Model and Health*, 15 J. HEALTH ECON. 49, 53-58 (1996).

86. See *infra* Section IV.B.

III. QALYS VERSUS WTP/WTAs AMOUNTS AS PROXIES FOR LWUs: THE ADVANTAGES OF QALYS

What are QALYs? What are WTP/WTa amounts? And how do they relate to LWUs, the welfarist gold standard?

Section A answers these questions. Sections B and C then demonstrate how QALYs can, under some conditions, be better proxies for LWUs than WTP/WTa amounts. Section B brackets issues of measurement and shows that QALYs, if accurately measured, can in some special contexts perfectly track LWUs, while accurately measured WTP/WTa amounts will not. Section C describes various cognitive phenomena that make the measurement of WTP/WTa amounts particularly difficult and that may not interfere as substantially with QALY measurement.

It should be stressed that Sections B and C are *not* meant to demonstrate the overall superiority of QALYs to WTP/WTa amounts as proxies for LWUs. QALYs have their own difficulties, which will be discussed in Part IV. But it is crucial to see that the comparison of QALYs and WTP/WTa is a mixed bag. WTP/WTAs are not perfect welfarist measures. They have disadvantages, as well as advantages, vis-a-vis other practicable measures such as QALYs. This point is wholly overlooked in the existing welfarist literature on QALYs, where the only reason advanced to prefer QALY-based policy analysis to WTP/WTa-based analysis is a taboo on monetization.⁸⁷

This Part therefore breaks new ground. Showing that WTP/WTAs do not dominate QALYs as a welfarist tool is a vital point, which policy-analytic scholarship has yet to grasp. My intention is not to ignore the limitations of QALYs, but rather to shine a bright light on the welfarist imperfections of the WTP/WTa measure, in this part of the Article, and then to do the same for QALYs in Part IV below.

A. What Are WTP/WTa Amounts? What Are QALYs?

QALYs and WTP/WTa amounts, like LWUs themselves, can be understood as idealized constructs. They constitute different ways to measure the welfare goodness of a life-history. These idealized constructs should serve to orient the actual practice of policy analysis: Responses to QALY surveys provide evidence that policy analysts can employ in estimating (idealized) QALYs, and similarly, responses to contingent-valuation surveys, or market behavior, are evidence of (idealized) WTP/WTa amounts.

Traditionally, cost-benefit analysis is defined as the sum of WTP/WTa: the

87. See *supra* Part I.

money amounts that individuals are willing to pay or accept in return for choices that affect them. Posner and I have suggested that WTP/WTa amounts may need to be “laundered” in various ways, to correct for perceptual and evaluative biases that may cause individuals to be mistaken about what truly benefits them.⁸⁸ This suggestion motivates the following definition of (idealized) WTP/WTa amounts:⁸⁹ If O is the status quo and O^* is an alternative outcome, then individual i ’s WTP/WTa for O^* is the amount of wealth subtracted from or added to her life-history L_i^* in O^* (at the present time) such that her life-history L_i^* , with this wealth subtracted or added, is just as good as her baseline life-history L_i in O . By “just as good,” I mean that the incremental resources just compensate for the difference in LWUs between L_i and L_i^* . The “laundering” here occurs because it is the preferences of idealized spectators—not the individual’s actual preferences—that determine the LWU values of the individual’s various possible lives.

What about QALYs? How do they fit into my framework of life-histories and LWUs? An individual’s longevity and morbidity profile is part of her life-history—part of the agglomeration of facts that affect her well-being—but not the only part. A given life history L_i might be thought of as a combination of a health-and-longevity history H_i , plus many other background features B_i , including individual consumption, recreational activities, social life, professional accomplishment, sexuality, and so on.⁹⁰ What do QALYs mean, given these other

88. See ADLER & POSNER, *COST-BENEFIT ANALYSIS*, *supra* note 22, ch. 5; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 220-22; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 22, at 289-300; Matthew Adler, *Incommensurability and Cost-Benefit Analysis*, 146 U. PA. L. REV. 1371, 1381-83 (1998) [hereinafter Adler, *Incommensurability and Cost-Benefit Analysis*].

89. In prior work, I have used the term “welfare equivalent” (WE) to mean an idealized or “laundered” WTP/WTa amount. So as to avoid burdening the reader with too much unfamiliar jargon, I will stick to the WTP/WTa terminology here. But the substance of my analysis is fully consistent with my prior claims that cost-benefit analysis, ideally, should be understood as the sum of WEs rather than the sum of actual WTP/WTa.

90. The philosophical literature on objective welfare goods specifies a range of plausible dimensions of human well-being, such as life, bodily health, bodily integrity, the use of the “senses, imagination and thought,” emotions, practical reason, affiliation, interaction with other species, play, and control over one’s environment (Martha Nussbaum’s list). See MARTHA C. NUSSBAUM, *WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH* 78-80 (2000); see also JOHN FINNIS, *NATURAL LAW AND NATURAL RIGHTS* 59-99 (1980) (identifying life itself, knowledge, play, aesthetic experience, sociability, practical reasonableness, and religion as objective welfare goods); JAMES GRIFFIN, *VALUE JUDGEMENT: IMPROVING OUR ETHICAL BELIEFS* 29-30 (1996) (identifying accomplishment, agency, understanding, enjoyment, and deep personal relations); DEREK PARFIT, *REASONS AND PERSONS* 499 (1984) (identifying moral goodness, rational activity, the development of one’s abilities, having children and being a good parent, knowledge, and the awareness of true beauty). These philosophers’ lists are, in effect, scholarly speculation about which feature of life-

features in the background? Note that this question about the interaction between QALYs and background characteristics does not arise for the extrawelfarist, for whom two individuals with identical health profiles, but different consumption, leisure, social, professional and sexual profiles, are equally “healthy.” The extrawelfarist social planner focused on the capability/resource/primary good of “health” can, quite correctly, insist that these lives should be assigned the same QALY value. But how can the welfarist social planner insist on that? Indeed, is it even a coherent enterprise for the welfarist to measure the welfare contribution of health to a life that depends not only on health, but much else?

In answering this question, I shall draw upon important recent scholarship in utility theory by Miyamoto, Wakker, Bleichrodt, and Peters.⁹¹ A very general problem in utility theory involves so-called “multidimensional” objects: objects with multiple types of attributes, falling into different domains or “dimensions.” The general question is whether the utility numbers for the objects can in turn be decomposed into a sum, a product, or some other relatively tractable function of “subutility” numbers representing the different dimensions.⁹² The simplest decomposition is additive, but it seems clear that the additive approach will not work for the problem at hand. A life-history L_i , as I have conceived it, is comprised of health-and-longevity attributes H_i plus background characteristics B_i . It would be nice if $W(L_i)$, the LWU measure of a given life-history, were representable as the sum of two subutility functions $Q(H_i) + V(B_i)$, but the requisite conditions for additive decomposition are too stringent to be met here.

histories idealized spectators would intrinsically prefer. See Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 297-300. The World Health Organization has recently undertaken a more rigorous and comprehensive process, involving focus groups, surveys, and expert consultation, to arrive at an index of quality of life, the World Health Organization Quality of Life (WHOQOL) assessment, that encompasses twenty-four attributes. Some of these involve health, but many do not (for example, positive feelings, self-esteem, body image, personal relationships, social support, sexual activity, home environment, financial resources, leisure, and spirituality). See The World Health Organization Quality of Life Group, *The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties*, 46 SOC. SCI. & MED. 1569, 1576-78 (1998) [hereinafter The WHOQOL Group]; *infra* text accompanying note 170 (discussing the WHOQOL).

In short, there is much disagreement about what the welfare-relevant dimensions of life-histories are, but no disagreement that there are a considerable number of dimensions and that welfare transcends mere health and longevity.

91. John M. Miyamoto et al., *The Zero-Condition: A Simplifying Assumption in QALY Measurement and Multiattribute Utility*, 44 MGMT. SCI. 839 (1998).

92. See generally FRENCH, *supra* note 77, at 102-201 (discussing multiattribute value theory and utility theory); RALPH L. KEENEY & HOWARD RAIFFA, DECISIONS WITH MULTIPLE OBJECTIVES: PREFERENCES AND VALUE TRADEOFFS (1976) (same); DETLOF VON WINTERFELDT & WARD EDWARDS, DECISION ANALYSIS AND BEHAVIORAL RESEARCH 259-350 (1986) (same).

Those conditions say that two dimensions are additively separable if and only if a given change in one dimension has the same effect on overall utility, regardless of where the object (here, life-history) is located in the other dimension.⁹³ That surely is not true of the interaction between health and longevity, on the one hand, and leisure, consumption, social interaction, professional accomplishment and the other non-health constituents of welfare, on the other. For example, it is typically and plausibly supposed by welfare economists that health can change the welfare benefit of consumption.⁹⁴ If that is true, then presumably health can also change the welfare benefit of sex, socializing and so on. An additive representation of lifetime welfare as the sum of health-and-longevity subutility plus background subutility would preclude all of these sorts of interactions.

What Miyamoto et al. demonstrate is that a *multiplicative* decomposition of utility is possible if two conditions are met: the “zero condition” and “standard-gamble invariance.” The “zero condition” says that one of the dimensions must have a zero level, such that all objects at that level in that dimension have equal utility regardless of where they are located in the other dimension. Standard-gamble invariance says that lotteries in which the first dimension (the one with the zero level) is varied and the second held fixed must be ranked the same way, regardless of where the second dimension is fixed.⁹⁵ These axioms, albeit technical and a bit difficult to grasp, turn out to have considerable relevance for QALYs, LWUs, and policy evaluation.⁹⁶ For I suggest that health-and-longevity

93. See, e.g., FRENCH, *supra* note 77, at 182-91; VON WINTERFELDT & EDWARDS, *supra* note 92, at 302-08, 334-41.

94. See, e.g., Evans & Viscusi, *supra* note 29, at 499-500; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 991-92 & n.11.

95. See Miyamoto et al., *supra* note 91, at 845-48. The Miyamoto conditions allow for any finite number of attributes or “dimensions.” I have stated the conditions as they apply to the two-attribute case, since I am here conceptualizing life-histories as involving two broad dimensions of welfare (each of course encompassing a variety of more specific subdimensions): health/longevity and background characteristics.

96. Han Bleichrodt and John Quiggin have also employed the standard-gamble invariance condition and zero condition to analyze the interaction between health and consumption, and James Hammitt has employed the standard-gamble invariance condition. See Han Bleichrodt & John Quiggin, *Life-Cycle Preferences over Consumption and Health: When Is Cost-Effectiveness Analysis Equivalent to Cost-Benefit Analysis?*, 18 J. HEALTH ECON. 681, 688-90 (1999); James Hammitt, *How Much Is a QALY Worth? Admissible Utility Functions for Health and Wealth* 3-6 (May 2002) (unpublished manuscript, prepared for Association of Environmental and Resources Economists at the Allied Social Sciences Associations meeting), available at www.feem.it/NR/Feem/resources/conferences/PRE2004-01-03-01.Hammitt.pdf [hereinafter Hammitt, *How Much Is a QALY Worth?*]. My analysis, with its focus on LWUs and the relative accuracy of QALYs and WTP/WTa in tracking LWUs, is quite different from Bleichrodt and Quiggin’s and Hammitt’s, but I am indebted to their work.

and background characteristics do interact in just the way required by the Miyamoto conditions. At a minimum, the assumption that they do thus interact is a reasonable working premise, one that I will consider relaxing later on but, for now will take as true.⁹⁷

The health-and-longevity dimension satisfies the “zero condition.” Because being alive is a precondition for consumption, leisure, sex, socializing, and so on, a life-history whose health profile H_i has zero longevity must be given the same ranking by extended spectators, and have the same LWU number, regardless of whatever background characteristics we might link to that life-history. As for standard-gamble invariance: Imagine that idealized spectators are indifferent between a given life history $L_i = (H_i, B_i)$, and a lottery over health characteristics holding fixed the background characteristics of that life-history, where there is a chance p of getting (H_i', B_i) and a chance $(1 - p)$ of getting (H_i'', B_i) . Then idealized spectators must also be indifferent between the same lottery over the health dimension with a different set of background characteristics. $L_i^* = (H_i, B_i^*)$ must be just as good as a chance p of getting (H_i', B_i^*) and a $(1 - p)$ chance of getting (H_i'', B_i^*) . Being wealthier or poorer, more or less social, more or less accomplished, cannot change the relative attractiveness of different health profiles and gambles over health profiles. This health-gamble-invariance condition, while surely less compelling than the proposition that all zero-longevity lives are equally good for welfare, seems just as plausible as standard restrictions on the structure of preferences that welfare economists employ so as to achieve tractable formalization and, therewith, cheap and feasible policy-evaluation procedures.⁹⁸

If spectators’ extended preferences over life-histories indeed satisfy the Miyamoto conditions, then $W(L_i) = Q(H_i) \times V(B_i)$. The welfare value of a life is representable as the multiplicative product of the value of its health-and-longevity characteristics and the value of its background characteristics. This brings us to QALYs. My proposal is that the QALY value of a given life be defined as this first value: $Q(H_i)$, the numerical value of the health-and-longevity profile partly constitutive of that life, which when multiplied by a numerical value $V(B_i)$ representing the consumption, leisure, professional, and other non-health attributes of the life, equals the overall utility or LWU value $W(L_i)$.

To reiterate: Assuming the Miyamoto conditions hold true, we can define a

97. See *infra* Section IV.A.

98. For example, it is standardly assumed that preferences are complete and transitive and often assumed that they can be represented by well-behaved (continuous, differentiable, monotonic, quasi-concave) functions. See, e.g., DAVID M. KREPS, A COURSE IN MICROECONOMIC THEORY 17-37 (1990) (discussing economic theory of consumer choice); CARL P. SIMON & LAWRENCE BLUME, MATHEMATICS FOR ECONOMISTS 544-57 (1994) (same).

“subutility” number $Q(H_i)$ encapsulating the health-and-longevity characteristics of a given life-history, and a “subutility” number $V(B_i)$ encapsulating the background characteristics of that life-history. The overall utility or LWU value of the life-history will be the multiplicative product of these two “subutilities.” H_i is a nonnumerical description of the nature and duration of the health states making up a life history. $Q(H_i)$ is the “subutility” that represents that health profile. Similarly, B_i is a nonnumerical description of the background characteristics of a life history. $V(B_i)$ is the subutility that represents that non-health profile. I define the QALY value of a life history as $Q(H_i)$. This QALY value is one input, along with $V(B_i)$, in a multiplicative formula that determines the LWU value associated with that life-history.

At this point, the reader might wonder how my definition of the QALY value of a life history, as $Q(H_i)$, relates to the standard formula for calculating QALYs, namely $\sum_j q(h_{i,j}) \times t_j$, where $h_{i,j}$ is the health of individual i in period j and t_j is the duration of period j . The answer is that $Q(H_i)$ is an idealized construct: the true health subutility of a life. $\sum_j q(h_{i,j}) \times t_j$ is a practicable formula for estimating $Q(H_i)$. If certain axioms hold true, the estimate will be perfect: $\sum_j q(h_{i,j}) \times t_j$ will necessarily equal $Q(H_i)$. If those axioms do not hold true, then $\sum_j q(h_{i,j}) \times t_j$ will not necessarily equal $Q(H_i)$, and must be understood as a rough estimate, perhaps quite rough. The axioms, and the possible divergence between the QALY value of a life and $\sum_j q(h_{i,j}) \times t_j$, are discussed below in Part IV.⁹⁹

B. How QALYs Can Improve on WTP/WTAs Amounts (Bracketing Measurement Problems)

This Section brackets measurement issues and shows that accurately measured QALYs, i.e., $Q(H_i)$ values, will, under certain conditions, be perfect proxies for LWUs. Under the same conditions, accurately measured WTP/WTAs amounts will not be perfect proxies for LWUs. This analysis is admittedly artificial, because policymakers do not have direct access to $Q(H_i)$ values, any more than they do to LWUs themselves. Measurement problems are in fact omnipresent. Still, the analysis will help highlight one of the potential advantages of nontraditional, QALY-based cost-benefit analysis, as compared to traditional, WTP/WTAs-based cost-benefit analysis.

The proposition that accurately measured QALYs will perfectly track LWUs, under some conditions, is a direct upshot of the Miyamoto conditions. The qualifying phrase “under some conditions” is crucial. If one life is different in the value of its background characteristics than another, then a health improvement in the first life will have a different impact on overall welfare than

99. See *infra* Section IV.B.

a health improvement in the second life. This is a point I will return to below.¹⁰⁰ But the reciprocal point is also true: Where the lives being compared have the same, or equally valuable, background characteristics, differences in their QALY values will perfectly track differences in their welfare goodness.

Imagine a class of life-histories $\{L_1, \dots, L_m\}$, composed of health profiles $\{H_1, \dots, H_m\}$ and profiles of consumption, leisure, and other background characteristics $\{B_1, \dots, B_m\}$. Assume that all the life-histories are equal in the value of these non-health profiles: that is, $V(B_1) = V(B_2) = \dots = V(B_m) = K$. As a shorthand, let us call this a “background-equivalent” class of lives. The QALY values of the lives may vary. $Q(H_1)$ does not necessarily equal $Q(H_2)$ or $Q(H_3)$ and so on. The Miyamoto model entails that $W(L_i) = Q(H_i) \times V(B_i)$, which reduces to $W(L_i) = Q(H_i) \times K$. In other words, the LWU value of a life within the background-equivalent class equals its QALY value multiplied by a constant K representing the background value of all the lives in the class. It follows that QALY changes are a perfect indicator of the welfare impacts of policy choices whose effects are limited to changing, either for certain or probabilistically, which life-histories within the class occur.

Consider first a policy that affects morbidity rather than mortality. In the status quo, individual 1 has life-history L_1 , and individual 2 has life-history L_2 . The policy would improve the first individual’s health, changing her life-history to L_1' , and degrade the second individual’s health, changing her life-history to L_2' . Assume, crucially, that the status quo and alternative possible lives, L_1 , L_2 , L_1' , and L_2' , fall within the same background-equivalent class. Then the overall change in QALYs achieved by implementing the policy would be $Q(H_1') + Q(H_2') - Q(H_1) - Q(H_2)$. And the overall change in welfare, represented by LWUs, would be $Q(H_1') \times K + Q(H_2') \times K - Q(H_1) \times K - Q(H_2) \times K = K[Q(H_1') + Q(H_2') - Q(H_1) - Q(H_2)]$. So the policy improves overall welfare if and only if it increases total QALYs.

Consider next a policy that affects mortality risks, again assuming that all the lives involved fall in the same background-equivalent class, with scaling constant K representing the value of the lives’ background characteristics. In the status quo, individual 1 has life-history L_1 , and individual 2 has life-history L_2 . The policy would create a risk p for individual 1 of life-history L_1' , and a risk $(1 - p)$ of life-history L_1'' , where L_1' and L_1'' involve a longer or shorter life-span than the baseline L_1 . Similarly, the second individual would incur a r risk of life-history L_2' , and a $(1 - r)$ risk of L_2'' , again with different life-spans than baseline L_2 . As before, each life history L_i has a component health history H_i ; in this example, the various health histories differ in the length of time that the subject lives. The expected QALY change from the policy is $p \times Q(H_1') + (1 - p) \times$

100. See *infra* Section IV.C.

$Q(H_1'') - Q(H_1) + r \times Q(H_2') + (1 - r) \times Q(H_2'') - Q(H_2)$. And the expected change in LWUs produced by the policy turns out to be that amount multiplied by the scaling constant K . Once more, the policy produces an expected increase in overall welfare if and only if it produces an expected increase in QALYs.

What about WTP/WTAs amounts? There are various ways in which WTP/WTAs measures of welfare effects on individual lives can fail to be perfectly correlated with utility measures of those effects. Our concern here is extended preferences and “laundered” or “idealized” WTP/WTAs amounts, but the imperfections in money measures of utility carry over to this context. WTP/WTAs amounts, again, are present wealth changes that compensate for welfare changes. Specifically, a present wealth increase compensates for a subject’s loss in welfare by increasing the subject’s expected material consumption over his lifetime. A present wealth reduction compensates for a subject’s increase in welfare by decreasing the subject’s expected material consumption over his lifetime.

One reason why WTP/WTAs amounts do not perfectly track LWUs is that changes in consumption need not translate into changes in LWUs at an interpersonally constant rate. To see this point, consider the following case. In the status quo, individual 1 has life L_1 ; if the policy were implemented, his life would instead be L_1' . In the status quo, individual 2 has life L_2 , replaced by L_2' under the policy. Each of these lives is a combination of health characteristics and background characteristics including consumption. So $L_1 = (H_1, B_1) = (H_1, (C_1, S_1, P_1 \dots))$; and $L_2 = (H_2, B_2) = (H_2, (C_2, S_2, P_2 \dots))$. The change in LWUs resulting from the policy would be $W(L_1') + W(L_2') - W(L_1) - W(L_2)$. Assume that present wealth changes directly translate into present changes in consumption. In other words, the first individual’s laundered WTP/WTA for the policy is that present change in his consumption profile $C_1', \Delta C_1'$, such that the lifetime welfare units of the amended L_1' equals $W(L_1)$. The same holds for the second individual’s WTP/WTA.

If LWU changes are proportional to WTP/WTAs amounts—that is, if $[W(L_1') - W(L_1)] / \Delta C_1' = [W(L_2') - W(L_2)] / \Delta C_2'$ —then the sum of WTP/WTAs does track LWUs. But there are a variety of reasons why, in general, the proportionality constraint just articulated need not hold true. Individual 1 may have a much higher level of lifetime consumption than individual 2, so that changes in his consumption have a smaller effect on lifetime welfare. Consumption has diminishing “marginal utility,” to use the standard economic lingo.¹⁰¹ Or sequencing effects might kick in: Although the two individuals have roughly the same level of lifetime consumption, the $\Delta C_1'$ would occur at a point in individual 1’s consumption sequence where his welfare is particularly

101. See, e.g., Adler, *Incommensurability and Cost-Benefit Analysis*, *supra* note 88, at 1398-1401.

sensitive, or insensitive, to consumption changes. Finally, consumption might interact with health, or with other background characteristics.¹⁰²

A different reason why the sum of WTP/WTAs amounts need not perfectly track the sum of LWUs is that present wealth changes need not always induce the same expected consumption changes in different individuals. WTP/WTAs amounts involve changes to present *wealth*, and the linkage between that change and the change in the individual's expected consumption (and therewith his welfare), need not be interpersonally constant. The so-called "dead anyway" effect, first discovered by Pratt and Zeckhauser,¹⁰³ and evident in various economic models of WTP/WTAs for longevity,¹⁰⁴ might be understood in these terms. To see the effect most simply, ignore non-consumption background characteristics, assume that the value of consumption is a constant function of lifetime consumption, and assume that all the individuals involved are healthy. So, for all lives, $W(L_i) = Q(H_i) \times V(C_i) = y_i \times k \times c_i$ (y_i is years, c_i is total lifetime consumption.) In the status quo, individual I is currently y_1 years old. He has a baseline risk r of dying immediately, in which case his lifetime consumption will have been c_1 , and $(1 - r)$ of living until age y_2 , in which case his lifetime consumption will be c_2 , which is larger than c_1 . He is asked for his willingness to pay for a policy that will reduce the risk of dying immediately to q , which is less than r .

The welfare change induced by the policy, measured in LWUs, turns out to be: $(r - q) \times k \times [y_2 c_2 - y_1 c_1]$. This is, as it should be, solely a function of the risk reduction $(r - q)$ and the difference in value of the lives involved ($k \times (y_2 c_2 - y_1 c_1)$). The LWU change is the same regardless of the absolute level of initial mortality risk r or residual risk q . But the amount of wealth—potential future consumption—that the individual is willing to sacrifice for the risk reduction depends on the absolute level of risk. In this case, $WTP = (r - q) \times (y_2 c_2 - y_1 c_1) / [(1 - q) \times y_2]$. So WTP is partly a function of q , the residual risk of dying that

102. See sources cited *supra* note 94.

103. See John W. Pratt & Richard Zeckhauser, *Willingness To Pay and the Distribution of Risk and Wealth*, 104 J. POL. ECON. 747, 750-53 (1996). For an accessible discussion of this effect, see Hammitt, *QALYs Versus WTP*, *supra* note 1, at 992-93.

104. See, e.g., Anna Alberini et al., *Does the Value of a Statistical Life Vary with Age and Health Status? Evidence from the US and Canada*, 48 J. ENVTL. ECON. & MGMT. 769, 771-73 (2004); Maureen Cropper & Frances Sussman, *Valuing Future Risks to Life*, 19 J. ENVTL. ECON. & MGMT. 160, 162-65 (1990); Hammitt, *QALYs Versus WTP*, *supra* note 1, at 992-93; cf. Friedrich Breyer & Stefan Felder, *The Dead-Anyway Effect Revisited* 9 (CESifo, Working Paper No. 805, 2002), available at <http://www.cesifo.de/~DocCIDL/805.pdf> (analyzing dead-anyway effect in models with and without bequests and under perfect and imperfect insurance markets and finding that "for individuals without a bequest motive the value of a statistical life always increases with the level of risk exposure if and only if they are risk-averse with respect to wealth").

remains after the risk reduction. Note that, the larger the residual risk q , holding constant the risk reduction ($r - q$), the larger WTP becomes. Why? The larger the residual risk, the larger the chance that the present resources that the individual is sacrificing will not eventuate in consumption: The individual will “die anyway.” The expected change in lifetime consumption for a present change in wealth depends, not just on the wealth change, but on the probability that death will intervene and preclude consumption. That is the essence of the “dead anyway” effect.¹⁰⁵

A different source of slippage in the wealth-consumption nexus is the fact that individuals might face different investment opportunities, for example investment horizons, so that (even bracketing differences in the risk of death) present wealth changes can produce different changes in lifetime consumption for different individuals.¹⁰⁶ Individuals 1 and 2 have identical health histories, leisure profiles, and so on, but the first is older than the second. At present, individual 1 will live for five more years, while individual 2 will live for twenty years. Both can invest their resources at a nonzero real interest rate, and spectators have no time preference for present over future consumption.¹⁰⁷ Then the two individuals will have different, “laundered” WTP/WTAs for the very same morbidity change. Specifically, the older individual will be willing to pay more to avoid increased morbidity, and demand more to accept it, because he has fewer years in which to invest present dollars in the market, so that present wealth increments translate into smaller consumption increments for him than for the younger individual. This “horizon” effect, along with the “dead anyway” effect, helps explain why the elderly might have inflated WTP/WTAs for morbidity or mortality changes, relative to the welfarist gold standard for measuring those changes, namely LWUs.

I have discussed different ways in which WTP/WTAs might be imperfect indicators of LWUs.¹⁰⁸ I have not yet shown that the WTP/WTAs can, under some conditions, be worse proxies than QALYs. To see that, consider the conditions where QALYs are perfect proxies: where all the lives affected by the policy choice fall in the same background-equivalent class. It is

105. See Olivier Armantier & Nicholas Treich, *Social Willingness To Pay, Mortality Risks and Contingent Valuation*, 29 J. RISK & UNCERTAINTY 7, 8 (2004) (showing that heterogeneity in individuals' survival probabilities can lead sum-of-WTP criterion to deviate from overall welfare).

106. This effect is evident in Yew-Kwang Ng's model of WTP/WTAs for longevity. See Yew-Kwang Ng, *The Older the More Valuable: Divergence Between Utility and Dollar Values of Life as One Ages*, 55 J. ECON. 1, 4-11 (1992).

107. See *infra* text accompanying notes 153-154 (suggesting that spectators might lack an intrinsic time preference).

108. Cf. Armantier & Treich, *supra* note 105, at 17 (using simulation analysis and finding that, where wealth, baseline risk, and risk reduction are heterogeneous and uncorrelated, the sum-of-WTP method overestimates overall welfare for a risk reduction project by fifteen percent).

straightforward to demonstrate that the various imperfections of WTP/WTa can occur under these conditions, and I do that in the margin.¹⁰⁹ To be sure, background equivalence is a special set of conditions; under other conditions, WTP/WTa may well track LWUs better than QALYs do. But my aim here is not to claim general superiority for QALYs. Rather, it is the more modest—but still important—aim of demonstrating that, contrary to the general wisdom in the welfarist literature on QALYs, QALYs are not dominated by WTP/WTa.

C. Measuring QALYs and WTP/WTa Amounts

The previous Section bracketed the problem of measuring QALYs and WTP/WTa and argued that a well-informed respondent's ("spectator's") QALY valuations for various life-histories, if perfectly measured, might under some conditions more accurately track her preferences over those possible lives than her WTP/WTa values, again perfectly measured. This Section suggests that QALY values might be easier to measure than WTP/WTa.¹¹⁰ Certain cognitive

109. Here are some simple models illustrating that WTP/WTa amounts can fail to be perfect proxies for LWUs even where all the lives affected by a policy choice fall in the same background-equivalent class. This is an important difference from QALYs, which *cannot* fail to be perfect proxies under these conditions. (1) *Diminishing marginal utility of consumption*. Imagine that each $L_i = (H_i, B_i)$, where background characteristics are decomposable into consumption C_i and other characteristics Z_i . By hypothesis, $W(L_i) = Q(H_i) \times V(B_i) = Q(H_i) \times V(C_i, Z_i)$. Assume that $V(C_i, Z_i)$ is increasing in C_i and Z_i ; that C_i has a diminishing marginal impact on $V(C_i, Z_i)$; and that the marginal impact of C_i does not decrease with increasing Z_i . Consider now a class of lives $\{L_i\}$ with the same V value = K . Some of these lives are high consumption, others low consumption, in each case with corresponding Z_i characteristics such that overall $V(C_i, Z_i) = K$. A larger ΔC_i will be required to compensate for a given health change (meaning here both the initial level of $Q(H_i)$ and the change in health value) and concomitant LWU change in a high-consumption life than in a low-consumption life. (2) *Interaction between consumption and health*. Assume that $L_i = (H_i, C_i)$ and that $W(L_i) = Q(H_i) \times c_i$. Consider the background equivalent class of lives with the same total consumption c . Imagine that the lives in the status quo are L_1 and L_2 and that a policy would improve their health to L_1' and L_2' , such that $Q(H_1') - Q(H_1) = Q(H_2') - Q(H_2) = q$ for both pairs of lives. It follows that $\Delta W = qc$ for both pairs of lives. But the Δc amount required to balance the change in the first life is $cq/Q(H_1')$, while the Δc amount required to balance the change in the second life is $cq/Q(H_2')$. (3) *Dead anyway effect*. Imagine healthy individuals who live y_i years, with consumption level c_i and further characteristics z_i such that $W(L_i) = y_i \times c_i \times z_i$. Within a background equivalent class, $c_i \times z_i = K$, or $z_i = K/c_i$. In the status quo, an individual has a risk r of dying now after living y_1 years and consuming c_1 , and risk $(1 - r)$ of living longer to y_2 years, consuming c_2 . He is asked for WTP to reduce r to q . The LWU change is $(r - q) \times (y_2 - y_1) \times K$. But the WTP amount equals $(r - q) \times (y_2 - y_1) \times c_2 / [y_2 \times (1 - q)]$.

110. But see Richard D. Smith, *The Relative Sensitivity of Willingness-To-Pay and Time-Trade-Off to Changes in Health Status: An Empirical Investigation*, 10 HEALTH ECON. 487, 495-96 (2001).

difficulties interfering with the elicitation of WTP/WTa for mortality and morbidity may not afflict QALY measurement, at least not as substantially.

In this Section, I will assume that the QALY formula $\sum_j q(h_{ij}) \times t_j$ is a reasonably good estimate of the true QALY value of a life, namely $Q(H_i)$. It may not be—as mentioned above, and further elaborated in the next Part. But if the axioms implying the equivalence of $Q(H_i)$ and $\sum_j q(h_{ij}) \times t_j$ are not grossly violated, then $Q(H_i)$ may well be easier to measure—via the additive formula $\sum_j q(h_{ij}) \times t_j$ —than WTP/WTa amounts.

Consider first the measurement of WTP/WTa amounts. One technique for determining these amounts is “contingent valuation”: an interview-based methodology. Various interviewing techniques have been developed.¹¹¹ Respondents might be asked to state the maximum amount of money they would be willing to pay in return for a good, or the minimum amount they would be willing to accept in exchange for a bad. Or they might be asked about their willingness to trade goods or bads for particular amounts of money, with the pattern of “yes” and “no” answers then used to infer a maximum WTP or minimum WTa. Whatever the technique employed, contingent-valuation studies in the area of health and mortality have collided with certain characteristic biases.

These biases emerge most dramatically in contingent-valuation studies of mortality or morbidity risks.¹¹² Expected utility theory implies that WTP for reducing the risk of death or harm should be nearly proportional to the change in risk, if it is small. “[I]f a reduction in annual mortality risk from 20 in 100,000 to 18 in 100,000 is valued at \$20, then a larger reduction from 20 to 16 in 100,000 should be valued at about \$40 (ignoring a tiny income effect).”¹¹³ Hammitt and

(finding WTP survey to be a more sensitive indicator of changes in health status than QALY survey).

111. For overviews of this methodology, see IAN BATEMAN ET AL., *ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES: A MANUAL* (2002); FREEMAN, *supra* note 56, at 161-87; VALUING ENVIRONMENTAL PREFERENCES: THEORY AND PRACTICE OF THE CONTINGENT VALUATION METHOD IN THE US, EU, AND DEVELOPING COUNTRIES (Ian J. Bateman & Kenneth G. Willis eds., 1999).

112. To be sure, the problems in measuring WTP/WTa for morbidity risks might be avoided by measuring WTP/WTa for disease states that occur with certainty, but the same is not true of mortality. WTa for certain death may well be infinite; WTP to avoid certain death may equal the subject’s entire stock of wealth. The money amounts that traditional cost-benefit analysts employ to value lifesaving, the so-called “value of statistical life” (VSL), are therefore derived from contingent-valuation or revealed-preference studies examining WTP/WTa for the risk of death. See Matthew D. Adler, *Against “Individual Risk”: A Sympathetic Critique of Risk Assessment*, 153 U. PA. L. REV. 1121, 1198 n.300 (2005) [hereinafter Adler, *Against “Individual Risk”*] (citing sources discussing the VSL method).

113. James K. Hammitt & John D. Graham, *Willingness To Pay for Health Protection: Inadequate Sensitivity to Probability?*, 8 J. RISK & UNCERTAINTY 33, 35 (1999).

Graham surveyed published contingent-valuation studies of mortality or health risks, focusing on the studies that allowed either “internal” tests of proportionality or “external” tests. If each respondent is asked for WTP for multiple risk reductions, and her responses are proportional, then that provides “internal” evidence of proportionality. Asking one group of respondents for their WTP for a particular risk reduction, and a different group for their WTP for a different reduction, permits an “external” test of proportionality.

Hammitt and Graham found that, of the ten studies permitting an “internal” test of proportionality, not a single study confirmed it.

[T]he average respondent [in these studies] does state a larger willingness to pay for larger risk reductions—i.e., the direction of change in payment size is in accordance with expectations. Yet a significant minority of respondents often report the same willingness to pay, regardless of the size of risk reduction . . . When the proportionality assumption is tested through internal tests, it generally fails. Mean willingness to pay is less—usually much less—than proportional to risk reduction. It is not uncommon for a doubling of risk reduction to be associated with far less than a 50% increase in payment.¹¹⁴

Proportionality was similarly disconfirmed by all nine of the “external” studies.¹¹⁵ Indeed, “[e]ven the less demanding test that willingness to pay be larger among respondents who are offered larger risk reductions, while satisfied in some studies . . . is not always satisfied in others.”¹¹⁶

What explains these disheartening results? In other contexts, for example contingent-valuation studies of environmental preservation, the insensitivity of WTP/WTa to the magnitude of the good produced or destroyed might reflect the fact that respondents are voicing moral rather than self-interested preferences.¹¹⁷ That is a less compelling explanation, however, of the magnitude-insensitivity that Hammitt and Graham describe.¹¹⁸ More plausibly, respondents queried about

114. *Id.* at 39; *see also id.* at 37-38 (listing the studies that the authors relied upon, showing that proportionality failed in every study where it could be tested).

115. *See id.* at 39-40.

116. *Id.* at 40. For similar findings that WTP/WTa is not sensitive to the magnitude of risk reduction, see Jane Beattie et al., *On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part I-Caveat Investigator*, 17 J. RISK & UNCERTAINTY 5 (1998), and the sources cited in U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 18; Jonathan Baron, *Biases in the Quantitative Measurement of Values for Public Decisions*, 122 PSYCHOL. BULL. 72, 74 (1997); and Hammitt, *QALYs Versus WTP*, *supra* note 1, at 997.

117. *See, e.g.,* Baron, *supra* note 116, at 75.

118. *See also* Alan Shiell & Lisa Gold, *Contingent Valuation in Health Care and the Persistence of Embedding Effects Without the Warm Glow*, 23 J. ECON. PSYCHOL. 251 (2002) (finding “embedding effects” in a health care contingent-valuation study designed to exclude moral

WTP/WTa for mortality and morbidity risks are affected by well-known biases that interfere with the evaluation of risk—for example, a misunderstanding of the basic rules of probability, or a departure from expected-utility theory in valuing lotteries, such as lotteries over personal health and longevity.¹¹⁹

A different set of biases, also presumably at play in the contingent-valuation studies about mortality and morbidity risk, involves tradeoffs between different kinds of goods—here money, on the one hand, and health and longevity, on the other. “[M]aking tradeoffs is a cognitively demanding task that people will try to minimize.”¹²⁰ There can also be emotional resistance to commensuration, for example where one good in the tradeoff is understood (pre-theoretically) by the respondent as “protected” relative to the second good—as qualitatively more valuable than the second.¹²¹ Life and health are often thought of this way, relative to money.¹²²

The literature on tradeoff difficulty suggests that respondents to contingent-valuation studies, facing cognitively or emotionally demanding questions about the relative significance of wealth and nonmonetary goods for their personal welfare, will refuse to answer entirely or (less dramatically) will employ various strategies for answering the questions—strategies that will produce stated WTP/WTa amounts that do not reflect their well-informed self-interested preferences for the goods. For example, the respondent may articulate a WTP/WTa amount that expresses the general “importance” of the type of good at issue, not the welfare-significance of the particular quantity involved. He may construct a mental budget that permits a limited expenditure on the good, and refuse to spend more than the budget even where additional expenditures would be worthwhile. Or the respondent may say what a fair price for the good would be, in effect focusing on the cost of supply rather than the personal benefits.¹²³

These sorts of tradeoff biases, together with risk biases, help explain the insensitivity of stated WTP to the magnitude of mortality or morbidity risk

preferences).

119. See PLOUS, *supra* note 78, at 84-188 (discussing these biases). *But see* Hammit & Graham, *supra* note 113, at 35 (“There is a variety of descriptive models of choice that predict [sic] responses to risk that are nonlinear in the probabilities [such as prospect theory] . . . but even these models are locally linear . . .”).

120. John W. Payne et al., *Measuring Constructed Preferences: Towards a Building Code*, 19 J. RISK & UNCERTAINTY 243, 257 (1999).

121. See *id.* at 257. On the cognitive and emotional obstacles to trading off different kinds of goods, see also MARY FRANCES LUCE ET AL., EMOTIONAL DECISIONS: TRADEOFF DIFFICULTY AND COPING IN CONSUMER CHOICE 1-10 (2001); and Baron, *supra* note 116, at 83-84.

122. See, e.g., Cass R. Sunstein, *Incommensurability and Valuation in Law*, 92 MICH. L. REV. 779, 834-40 (1994).

123. See Baron, *supra* note 116, at 75-77, 79-80, 83-84; Shiell & Gold, *supra* note 118, at 258-60.

reduction. They also imply that stated WTP may inaccurately measure health benefits even where risk processing is not an issue. That implication seems to be confirmed by a study conducted by Alan Shiell and Lisa Gold, who asked respondents for WTP for two vaccines to treat two different infectious diseases and for a composite vaccine for both diseases.¹²⁴ In a majority of cases, the respondent's WTP for the composite vaccine was less than the sum of WTP for the two individual vaccines. Similar evidence of magnitude-insensitivity apart from risk¹²⁵ shows up in other health contingent-valuation studies.¹²⁶

My discussion thus far has focused on the contingent-valuation format and the evidence that risk and tradeoff biases interfere with the measurement of WTP/WTa for mortality and morbidity in that context. To be sure, a variety of so-called "revealed preference," noninterview techniques for estimating WTP/WTa are also widely used by applied economists.¹²⁷ Here the idea is to infer valuation from behavior. There is a large empirical literature that seeks to correlate wage differences with differences in occupational risks.¹²⁸ A smaller literature uses nonoccupational data to estimate WTP/WTa for mortality risks (for example, consumer purchases of safety devices), or morbidity (for example, using the variation in housing prices to estimate WTP/WTa for air pollution and related morbidity).¹²⁹ Revealed-preference techniques may eliminate some biases associated with the contingent-valuation method—for example, the so-called "hypothetical bias," namely that asking respondents what they would be willing to pay or accept in hypothetical choice situations may not elicit sincere and considered statements of the respondents' real preferences¹³⁰—but it is hard to see how the shift from discursive to observational preference-measurement methods would eliminate risk or tradeoff biases. Individuals who process

124. See Shiell & Gold, *supra* note 118, at 253-58.

125. To be sure, risk comes into play in the Shiell and Gold study because the vaccinations would merely reduce the risk of the infectious diseases, but given the study design it is hard to see how difficulties in processing risk would fully explain the magnitude-insensitivity observed.

126. See Jan Abel Olsen et al., *The Insensitivity of "Willingness-To-Pay" to the Size of the Good: New Evidence for Health Care*, 25 J. ECON. PSYCHOL. 445 (2004); *id.* at 447 (citing literature); Shiell & Gold, *supra* note 118, at 258-59 (same).

127. For an overview of these techniques, see FREEMAN, *supra* note 56, at 95-136.

128. See, e.g., *id.* at 317-19, 401-06; VISCUSI, *supra* note 56, at 34-65.

129. These literatures are discussed in VISCUSI, *supra* note 56, at 65-67; F. Reed Johnson et al., *Valuing Morbidity: An Integration of the Willingness-To-Pay and Health-Status Index Literatures*, 16 J. HEALTH ECON. 641, 644 (1997); and Richard Clemmer et al., *Household Health Production, Property Values, and the Value of Health*, in VALUING HEALTH FOR POLICY, *supra* note 1, at 105.

130. See, e.g., Kevin J. Boyle & John C. Bergstrom, *Doubt, Doubts, and Doubters: The Genesis of a New Research Agenda?*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 111, at 184-86 (discussing concerns about the hypothetical nature of contingent-valuation questions).

probabilities irrationally, or have cognitive or emotional difficulties trading life or limb for money, should exhibit insensitivity to the magnitude of risk or morbidity reduction in their purchasing or precautionary behavior.¹³¹

What about QALYs? My suggestion is that risk and tradeoff biases may pose less of a problem for the measurement of QALY values—assuming $\sum_j q(h_{i,j}) \times t_j$ is a good estimate of the genuine QALY value of a life, i.e., $Q(H_i)$ —than for the measurement of WTP/WTa. Consider first the QALY measurement of pure mortality or mortality risk—that is, measuring the QALY loss for each affected individual of a policy which causes each to lose ΔT_i years of healthy life or (more realistically) imposes an incremental r_i risk of losing ΔT_i years of healthy life. In these cases, proportionality to the amount of longevity at stake (ΔT_i) and the risk (r_i) is guaranteed by the additive formula $\sum_j q(h_{i,j}) \times t_j$. If $Q(H_i) = \sum_j q(h_{i,j}) \times t_j$ then, in the case of an individual who lives T_i years in perfect health, the QALY value of her life-history is simply T_i . If she loses or gains ΔT_i years, the QALY measure of the change is ΔT_i . If healthy individual 1 loses ΔT_1 years, and healthy individual 2 loses ΔT_2 years, and ΔT_2 is twice ΔT_1 , then the QALY value of the second loss is automatically twice the QALY measure of the first. Tradeoff, risk and other biases are, in this limiting case, wholly circumvented because QALY values are derived using the additive formula, not from individual statements of preference. Ditto for the case in which individual 1 incurs an r_1 risk of losing ΔT_1 years of life and individual 2 incurs an r_2 risk of losing ΔT_2 years. The ratio of the expected QALY losses for the 2 individuals is necessarily equal to the ratio of the longevity risks that the individuals face ($(r_2 \times \Delta T_2) / (r_1 \times \Delta T_1)$), regardless of what the individuals might say about the dollar or nondollar value of these risks.

Of course, the QALY method is a tool for measuring morbidity and longevity on an integrated scale, and individual statements of preference are crucial in determining morbidity values—thus the massive efforts undertaken by QALY researchers to survey doctors, patients, or members of the public about their ranking of health states. But even where morbidity comes into play, biases that might induce individual insensitivity to the magnitude of morbidity change are still partly avoided through the additive formula $\sum_j q(h_{i,j}) \times t_j$. Consider a policymaker faced with the options of a temporary or medium-term reduction of

131. I am aware of no study that tests magnitude sensitivity in the revealed preference context. It may be difficult to design such a study. See E-mail from James K. Hammitt, Professor of Economics and Decision Sciences, Harvard School of Public Health, to Matthew D. Adler, Professor of Law, University of Pennsylvania Law School (Oct. 29, 2004, 11:38:53 EST) (on file with author). For a general discussion of difficulties with revealed-preference studies of WTP/WTa for morbidity and mortality, see U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 18-19; Johnson et al., *supra* note 129, at 644; Hammitt & Graham, *supra* note 113, at 33-34; and Hammitt, *QALYs Versus WTP*, *supra* note 1, at 997-98.

some form of air pollution, thereby producing a temporary (one-year, say) or medium-term (five-year) abatement of asthma among certain asthmatics whose symptoms are caused by the pollution. Assume, for the sake of illustration, that the pollution causes asthma only on days when weather patterns lead to a particularly high amount of inhaled pollutant, which occur on average every tenth day. Asthma has a QALY value of 0.683.¹³² Then the per-individual QALY value of the medium-term reduction ($0.317 \times 5 \times 0.1 = 0.1585$ QALY) will automatically be five times the per-individual value of the temporary reduction ($0.317 \times 1 \times 0.1 = 0.0317$ QALY). Bracketing income effects, WTP for the medium-term reduction should also be five times WTP for the temporary reduction; but in fact contingent-valuation studies eliciting WTP for longer versus shorter relief from the same symptoms generally fall far short of proportionality.¹³³

As for the measurement of morbidity values themselves: Risk and tradeoff biases will, to some extent, affect these measurements. The two most widely accepted methods for eliciting morbidity values, as already mentioned, are the standard-gamble (SG) and time-tradeoff (TTO) methods.¹³⁴ The SG format, again, seeks to determine respondent's indifference probability p between a given health state, and a lottery with probability p of perfect health and $(1 - p)$ of dying immediately; while the TTO format focuses on the indifference ratio x/y between living for y years in the health state, and x years in perfect health. There is a substantial meta-literature that looks at the practicability, reliability, and validity

132. See Johnson et al., *supra* note 129, at 651.

133. See Johnson et al., *supra* note 129, at 650-51 (summarizing results of contingent-valuation studies of short-term health conditions); Thomas Klose, *The Contingent Valuation Method in Health Care*, 47 HEALTH POL'Y 97, 106 (1999) (discussing magnitude insensitivity in these studies); U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 25 (same). But see Richard M. O'Connor et al., *Urge Incontinence: Quality of Life and Patients' Valuation of Symptom Reduction*, 14 PHARMACOECONOMICS 531, 536-37 (1998) (finding that WTP is sensitive to the degree of reduction of urinary incontinence). For a recent survey of the health care contingent valuation literature, see Jan Abel Olsen & Richard D. Smith, *Theory Versus Practice: A Review of 'Willingness-To-Pay' in Health and Health Care*, 10 HEALTH ECON. 39 (2001).

134. Good reviews of the SG, TTO, and other methods employed to elicit QALY valuations of morbidity are: U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 16-18; Baron, *supra* note 116, at 80-82; Brazier et al., *supra* note 1, at 23-56; Dolan, *supra* note 1, at 1732-36; and Johannesson et al., *supra* note 1, at 283-85. The visual analogue scale ("VAS") approach to eliciting QALYs, like SG and TTO, is widely used, see Brazier, *supra*, at 24, but it is quite controversial, because of concerns that VAS valuations have no theoretical foundations and are merely ordinal, rather than cardinal measures, see Brazier, *supra* at 34-35; Dolan, *supra*, at 1733. Two other methods, magnitude estimation and the person-tradeoff, are less often used. See Brazier, *supra*, at 24-27.

of the two methods.¹³⁵ The methods appear to be practicable as well as “internally” reliable and valid, in the sense that a high proportion of respondents complete the surveys with internally consistent values that are stable over time. Elicited values diverge, to some extent, among respondents,¹³⁶ but this itself does not impugn the QALY technique, since (as I have already discussed) convergence in preferences or extended preferences is a contingent matter.¹³⁷ Other questions about the validity of SG and TTO values have to do with the decomposability of preferences for lifetime health histories into the sum of the durations of the component health states weighted by valuations of those states—an issue I will consider below.¹³⁸ Here, the problem is not really that biases are interfering with the measurement of SG and TTO values, but rather that the equivalence of $\sum_j q(h_{i,j}) \times t_j$ and $Q(H_i)$ presupposed by techniques for eliciting SG and TTO values is breaking down.

That said, there is some evidence that biases can interfere with the measurement of SG and TTO values.¹³⁹ The SG format involves both risk (since respondents are asked for indifference probabilities) and tradeoffs among different dimensions (since a lottery with a chance of dying is being compared to a health state). Note, however, that the different dimensions are health and longevity, not health or longevity and money. So the problem of trading off “protected” for “unprotected” goods is less pressing—certainly when it comes to trading off longevity for improvements in grave health conditions. Note also that risk biases can be eliminated by shifting to the TTO format.¹⁴⁰ Tradeoff biases remain, here,¹⁴¹ but respondents do generally voice TTO values that are less than one—indicating that they have traded off some longevity for an improvement in the health state—and voice lower TTO values for more serious conditions. Dolan, in a very large TTO survey of the general public (3395 respondents) in the United Kingdom, found that:

135. The meta-literature is exhaustively reviewed in Brazier et al., *supra* note 1, at 30-34, 36-39.

136. See Dolan, *supra* note 6, at 150.

137. See *supra* text accompanying notes 73-75. To be sure, the fact that some divergence in QALY values is to be expected, given the possibility that different individuals might have divergent preferences over life-histories, is hardly the end of the story. When survey respondents voice divergent QALY values, which ones should policymakers use? This is an important but difficult question which, along with the related problem of divergent extended preferences, I will not attempt to address in this Article.

138. See *infra* Section IV.B.

139. See Brazier et al., *supra* note 1, at 30-34, 36-39.

140. It should be emphasized that, *if* the conditions for additive decomposition of preferences over lifetime health histories as per the QALY formula obtain, *then* the TTO and SG methods should produce identical valuations of health states. See Johannesson et al., *supra* note 1, at 285; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 995.

141. See Brazier, *supra* note 1, at 38-39.

46% of respondents were willing to sacrifice life expectancy to avoid *all* of the [13] dysfunctional states they were presented with A further 29% were willing to sacrifice life expectancy for all but one or two of the states. In such cases, the unwillingness to trade-off time was almost exclusively associated with one or both of the very mild states. In all, 95% of respondents were prepared to sacrifice life expectancy for 6 or more states.¹⁴²

In a different, smaller survey, Dolan looked at each respondent's "consistency rate"—the percentage of pairings of more and less serious conditions in which the respondent gave the more serious condition a lower TTO value—and found that the median respondent had a consistency rate exceeding ninety percent.¹⁴³

IV. THE LIMITATIONS OF QALYS AS PROXIES FOR LWUS

Although QALYs have certain advantages over WTP/WTAs amounts as proxies for LWUs, QALYs are hardly a perfect welfarist measuring rod. This Part discusses their limitations, in particular: violations of the basic presupposition that the health and non-health characteristics of welfare are multiplicatively separable, such that $W(L_i) = Q(H_i) \times V(B_i)$; violations of the further conditions requisite for an additive decomposition of the overall QALY value of a health-profile, $Q(H_i)$, into the sum of the quality-weighted durations of its component health states; the dependence of LWUs on both QALYs and background characteristics; and the related fact that QALYs cannot be used to measure changes in the non-health determinants of well-being, by contrast with the WTP/WTAs method, which in principle is applicable to all aspects of welfare.

A. Is Lifetime Utility Separable into Health-Related Subutility and Background Subutility?

QALY values are typically elicited, in QALY surveys, without any discussion of the background characteristics (wealth or other attributes) that the subjects whose health states are being valued should be assumed to have.¹⁴⁴ This procedure is justified, for the welfarist, only if the Miyamoto "standard-gamble invariance" condition is satisfied: Namely, respondents (or at least idealized spectators) should have preferences among lotteries over different lifetime health

142. Dolan, *supra* note 6, at 149.

143. See P. Dolan et al., *Valuing Health States: A Comparison of Methods*, 15 J. HEALTH ECON. 209, 217-20 (1996).

144. See Hammitt, *How Much Is a QALY Worth?*, *supra* note 96, at 5; Mark J. Sculpher & Bernie J. O'Brien, *Income Effects of Reduced Health and Health Effects of Reduced Income: Implications for Health-State Valuation*, 20 MED. DECISION MAKING 207, 209, 211 (2000).

histories that are the same for every fixed level of background characteristics. If this condition holds true, along with the “zero condition” (less contestable), then respondents’ utilities for life-histories are expressible as the multiplicative product of a health subutility and a background subutility, and stated QALY values should be the same regardless of the background characteristics, as long as they are held fixed.¹⁴⁵ If the Miyamoto “standard-gamble invariance” condition does not hold true, then the current elicitation procedures are problematic. More fundamentally, if $W(L_i)$ is not equal to $Q(H_i) \times V(B_i)$, it is not clear what the QALY value of a life-history means for the welfarist. At a minimum, if the Miyamoto condition fails and $W(L_i)$ does not equal $Q(H_i) \times V(B_i)$, the stated valuations that respondents provide in QALY surveys will be less useful to policymakers as proxies for LWUs.

Keeny and Raiffa, in their seminal work on multiattribute utility theory, write that different attributes satisfy “independence” conditions such as standard-gamble invariance “[i]n a surprisingly large number of contexts.”¹⁴⁶ I have suggested that, intuitively, the Miyamoto condition is at least approximately true of health versus non-health attributes. That suggestion is, in effect, just a guess about what spectators’ preferences would be. The guess is not particularly well-informed. There appears to be very little empirical scholarship, at least in the QALY field, examining the structure of preferences over combinations of health profiles and background profiles.¹⁴⁷

QALY scholars have examined a related issue: What predicts variations among respondents in QALY values? “[M]ost studies show[] that the values are independent from socio-economic factors or professional level.”¹⁴⁸ This finding might be adduced as evidence for the standard-gamble invariance condition. If (1) preferences for lotteries over health characteristics do depend on the level of background characteristics, and (2) respondents to QALY surveys assume, absent instruction, that the health states being valued are packaged with the respondents’ actual background characteristics, it would follow that (3) variation in QALY values would correlate with variation in respondents’ background characteristics—which has not been observed. However, the finding of non-correlation is at best circumstantial evidence for the standard-gamble invariance

145. See *supra* text accompanying notes 91-98.

146. KEENY & RAIFFA, *supra* note 92, at 226.

147. See Hammitt, *How Much Is a QALY Worth?*, *supra* note 96, at 5 (“The literature on QALYs is virtually silent on the extent to which [valuations of health states] depend[] on wealth, income, or consumption.”).

148. U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 21; see also Dolan, *supra* note 1, at 1747; Paul Dolan & Jennifer Roberts, *To What Extent Can We Explain Time Trade-Off Values from Other Information About Respondents?*, 54 SOC. SCI. & MED. 919, 927-28 (2002).

condition.¹⁴⁹ More direct testing remains to be done.¹⁵⁰

B. Can QALYs Be Decomposed as per the Standard Additive Formula?

A different issue concerns the additive decomposition of $Q(H_i)$ into $\sum_j q(h_{i,j}) \times t_j$, where $h_{i,j}$ is the health of individual i in period j and t_j is the duration of period j . The vast bulk of the QALY literature assumes that the health value of a lifetime health-history or “profile” can, at least approximately, be represented as the sum of years in its component health states adjusted by values for those health states.¹⁵¹ Cost-effectiveness studies use this additive formula for determining the effectiveness of different interventions; and typical QALY surveys, which attempt to elicit SG or TTO values for states rather than whole histories, assume additivity as well. But additivity is *not* entailed by the more basic conditions—the Miyamoto conditions—that allow the LWU assigned to each life-history to be represented as the product of an overall health subutility for its component health history, $Q(H_i)$, and an overall non-health subutility for its component background characteristics, $V(B_i)$. Rather, theoretical work on QALYs has demonstrated that three further conditions that are “internal” to the valuation of health are required for this overall health utility $Q(H_i)$ to be additively decomposable into $\sum_j q(h_{i,j}) \times t_j$: (1) No discounting of future health; (2) risk neutrality with respect to longevity; and (3) no sequencing effects.¹⁵²

The no-discounting condition is the least problematic. No discounting means that the spectator’s present valuation of a life history in which he incurs a given health state does not vary depending on whether that state occurs in the near or more distant future. Although actual individuals often exhibit a time

149. For example, it may be that respondents assume that the subject’s background characteristics are population-average characteristics, rather than the respondents’ own characteristics. In this case we would not expect QALY values to vary depending with respondents’ background characteristics even if the Miyamoto condition fails.

150. For an unusual example of a direct test of the independence of valuations of health and background characteristics, see Antonio Ciampi et al., *Measurement of Individual Preferences: The Importance of “Situation-Specific” Variables*, 2 MED. DECISION MAKING 483 (1982).

151. To be sure, the HYE approach drops this assumption; but HYE’s have not been used much in practice, see Hammitt, *QALYs Versus WTP*, *supra* note 1, at 989.

152. For a discussion of these conditions or equivalent ones, see generally Han Bleichrodt, *QALYs and HYE’s: Under What Conditions Are They Equivalent?*, 14 J. HEALTH ECON. 17, 20-25 (1995); Han Bleichrodt et al., *Characterizing QALYs by Risk Neutrality*, 15 J. RISK & UNCERTAINTY 107 (1997); Dolan, *supra* note 1, at 1729-31, 1740-43; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 986-88; Magnus Johannesson et al., *A Note on QALYs, Time Tradeoff, and Discounting*, 14 MED. DECISION MAKING 188 (1994); Johannesson et al., *supra* note 1, at 285-86; Miyamoto et al., *supra* note 91, at 839-45; and Joseph S. Pliskin et al., *Utility Functions for Life Years and Health Status*, 28 OPERATIONS RES. 206, 207-15 (1980).

preference,¹⁵³ preferring present to future pleasures and future to present pains, and although policymakers certainly might use discounting to reflect the opportunity cost of current expenditures given the alternative of investing the resources in inter-temporal markets, a strong case can be made that the absence of an intrinsic time preference is normative.¹⁵⁴ Idealized spectators, one might suppose, would not give greater weight to the temporally proximate aspects of the lives they might lead just because of the proximity. In any event, if the risk-neutrality and no-sequencing conditions hold true, the formula $\sum_j q(h_{i,j}) \times t_j$ can be adjusted to incorporate a temporal discount factor.¹⁵⁵

Risk-neutrality and no sequencing are more serious difficulties for this formula. The risk-neutrality condition, here, means that longevity has constant rather than increasing or decreasing marginal health utility. Formally, for any health profile consisting of a constant health state, risk neutrality requires that a given increment in longevity produce the same increase in the overall $Q(H_i)$ value of the health profile regardless of the baseline longevity. For example, doubling the lifespan of an individual with chronic bronchitis from thirty-five to seventy years doubles the $Q(H_i)$ value. It is obvious that risk-neutrality in this sense is presupposed by the $\sum_j q(h_{i,j}) \times t_j$ formula, which makes estimated QALY values a linear function of longevity, holding health constant. But $Q(H_i)$ need not be a linear function of longevity, holding health constant, if the idealized spectators whose preferences ground LWU and $Q(H_i)$ values depart from risk-neutrality.

Would they? Empirical tests find that actual respondents are not always risk-neutral with respect to longevity.¹⁵⁶ Nor can it be said here—by contrast with discounting—that risk-neutrality with respect to longevity is normative. If money can have a declining marginal impact on welfare, then presumably so can lifespan. One solution, proposed in the literature, is to adjust the longevity component of a health history by a risk-aversion factor. In the case of a chronic health state h_i^* , $Q(H_i)$ would equal the value of that state $q(h_i^*)$ multiplied by the

153. See, e.g., Bleichrodt et al., *supra* note 152, at 110. Surprisingly, actual individuals may also exhibit a negative time preference. See Dolan, *supra* note 1, at 1742.

154. See David O. Brink, *Prudence and Authenticity: Intrapersonal Conflicts of Value*, 112 PHIL. REV. 215 (2003).

155. See, e.g., Johannesson et al., *A Note on QALYs*, *supra* note 152. Discounting is not consistent with risk neutrality over life years, but it is consistent with risk neutrality over discounted life years. See *id.*

156. Risk neutrality can be tested directly (by asking for the number of years in a health state that the respondent views as equivalent to a gamble over the chance of immediate death and longer duration in the health state) or indirectly, by testing the other conditions that are implied by risk neutrality plus the zero condition. See Bleichrodt et al., *supra* note 152, at 112-13. Both sorts of test show that risk neutrality can fail. See Dolan, *supra* note 1, at 1740-42.

longevity raised to a risk-aversion factor.¹⁵⁷ But it is unclear how one might extend this formula to accommodate risk aversion in the more realistic case of health profiles where health varies over time.¹⁵⁸ Further, even in the case of unvarying health profiles, there are empirically documented instances of preferences that cannot be represented by either the ordinary QALY formula *or* the risk-adjusted version: namely the preference to live more than zero time, but less than an unbearable amount of time, in a painful health state.¹⁵⁹ Once more, it is hard to see why general considerations of rationality would preclude spectators from having such preferences. In short, departures from risk neutrality can create a divergence between the $\sum_j q(h_{i,j}) \times t_j$ formula and the genuine QALY value of a life $Q(H_i)$, and there appear to be no simple modifications to the QALY formula to circumvent the problem.

The same is true of sequencing. In a few studies, researchers have tested the no-sequencing condition directly by asking respondents to value sequences of states, and comparing the values with the sum of QALY values for the states involved.¹⁶⁰ Other research employs more indirect tests.¹⁶¹ The literature is small and conclusions are mixed, but sequencing does emerge in some cases. For example, Richardson asked women to value separately three differentially serious breast cancer states, then a deteriorating holistic scenario composed of a progression from less to more serious, and found that the values could not be reconciled even allowing for discounting. A better explanation, he suggests, is that the “knowledge of future suffering and death casts a shadow over, or devalues, the enjoyment of earlier life years.”¹⁶² Note how this coheres with the philosopher David Velleman’s suggestion that a life history where momentary welfare has a deteriorating trajectory is worse than a counterpart life history with the same overall sum of momentary welfare but an increasing trajectory.¹⁶³

C. Background Characteristics Redux

Even if LWUs can be decomposed into the product of health and non-health subutility, and even if health subutility can in turn be decomposed as per the

157. See Johannesson et al., *supra* note 152, at 188-90; Miyamoto et al., *supra* note 91, at 842-45.

158. See Johannesson et al., *supra* note 1, at 285.

159. See, e.g., Miyamoto et al., *supra* note 91, at 844-45; Paul Dolan & Peep Stalmeier, *The Validity of Time Trade-Off Values in Calculating QALYs: Constant Proportional Time Trade-Off Versus the Proportional Heuristic*, 22 J. HEALTH ECON. 445 (2003).

160. See Kuppermann et al., *supra* note 69; Richardson et al., *supra* note 69.

161. See Krabbe et al., *supra* note 69; Spencer, *supra* note 69; Treadwell, *supra* note 69.

162. Richardson et al., *supra* note 69, at 157.

163. See *supra* text accompanying note 84.

additive QALY formula into the cumulative time in different health states adjusted for the value of those states, QALYs are not generally a perfect proxy for LWUs. The problem of background characteristics reemerges here. I demonstrated earlier that QALY aggregation does perfectly track overall well-being where all the lives affected by the policy choices being considered fall in the same background-equivalent class. In general, of course, that need not be true. If life history L_1 has a higher level of consumption, or leisure, or socializing, or professional accomplishment, as compared to L_2 , then a change in L_1 's health history has a greater impact on overall well-being than the very same change (as measured in QALY units) in L_2 's health history. This follows immediately from the multiplicative representation of LWUs implied by the Miyamoto conditions, i.e., the standard-gamble invariance and zero conditions.¹⁶⁴ If those conditions obtain and the multiplicative model therefore does indeed accurately represent LWUs, health *cannot* generally have equal welfare value for different persons; rather, its value is necessarily scaled up or down by the value of the non-health attributes, $V(B_i)$, with which health partly interacts.

Rejecting the standard-gamble invariance condition, which would mean that $W(L_i)$ need not equal $Q(H_i) \times V(B_i)$, will not solve the problem. In that event, LWUs will neither be additively nor multiplicatively separable into health and non-health subutilities, and health and non-health attributes would continue to interact, albeit in a more complicated way. Rejecting the zero condition and adding other conditions to ensure the additive separability of health and non-health attributes *would* ensure that equal QALYs have equal welfare value across persons. An additive decomposition, here, would mean that $W(L_i) = Q(H_i) + V(B_i)$, precluding any interaction between health and non-health attributes. But it is implausible, given our sense of what well-informed individuals can prefer, that spectators' preferences would separate health and background characteristics so completely. The additive form would mean that the marginal utility of consumption cannot depend on health, nor vice versa.¹⁶⁵

Within the context of the multiplicative model, $W(L_i) = Q(H_i) \times V(B_i)$, is there any way to limit the scaling effect and make QALYs a reasonable proxy for

164. Bleichrodt and Quiggin make a similar point in the context of their own model, which (like mine) has a multiplicative structure deriving from standard-gamble invariance and the zero condition:

[In our model] the utility of health status is multiplied by the utility of consumption. Consequently, a given gain in quality of life will be more appreciated at higher levels of consumption. This implies that in the allocation of health care resources, larger welfare gains can be obtained by devoting resources to those individuals who have a high level of general consumption. . . . This result is ethically troubling However, the need for a multiplicative utility structure shows that . . . such implications cannot be escaped.

Bleichrodt & Quiggin, *supra* note 96, at 685.

165. See *supra* text accompanying notes 93-94 (discussing additive decomposition).

LWUs even where all the lives involved do not fall in the same background-equivalence class? One way to do that might be to expand the definition of health. If health merely subsumes the physical condition of the subject, excluding his hedonic, emotional, or cognitive state, then hedonic, emotional, and cognitive attributes become background characteristics. QALYs, as a measure of health thus narrowly defined, would be a poorer proxy for welfare than if health were defined more inclusively—since a narrower definition of health, or equivalently a broader definition of background characteristics, implies a smaller number of choice situations in which the lives involved fall in the same background-equivalent class, and presumably a greater average range of the scaling factor $V(B_i)$ in other choice situations.

Fortunately, health for QALY purposes is not normally defined so narrowly. QALY surveys are often conducted using so-called “health state classification systems,” which seek to regiment the evaluation task by describing health states as packages of health attributes.¹⁶⁶ Respondents are then asked to use the SG, TTO, or some other method to place the packages on a 0-1 scale, with 0 meaning death and 1 meaning the very best package. For example, the Health Utilities Index, one of the most widely used health state classification systems,¹⁶⁷ conceptualizes health states as a combination of vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Each of these eight attributes has five or six levels. The very best health state in the HUI system, i.e., the state with the QALY value of 1, is a state where the subject is at the best level with respect to all eight attributes. He is able to see well enough to read newsprint and recognize a friend across the street without glasses (vision); is able to hear what is said in a group conversation with at least three other people, without a hearing aid (hearing); is able to be understood completely when speaking with strangers or friends (speech); is able to walk around the neighborhood without difficulty and without walking equipment (ambulation); has the full use of two hands and ten fingers (dexterity); is happy and interested in life (emotion); is able to remember most things, think clearly, and solve day-to-day problems (cognition); and is free of pain and discomfort (pain).¹⁶⁸ The crucial point, for our purposes, is

166. See Dolan, *supra* note 1, at 1731-32, 1744-45 (discussing generally health state classification systems); QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 161-362 (describing specific systems in detail).

167. See Chaim Bell et al., *An Off-the-Shelf Help List: A Comprehensive Catalog of Preference Scores from Published Cost-Utility Analyses*, 21 MED. DECISION MAKING 288, 290 (2001) (finding that the Rosser Index, Quality of Well-Being Scale, and HUI are the most widely used health-state classification systems in published cost-utility analyses).

168. See David H. Feeny et al., *Health Utilities Index*, in QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 239 (describing the HUI system). The HUI system has been updated several times, and the version described in the text is the most recent one.

that the HUI system does have a more inclusive definition of health than the merely physical; and this is true of the other health classification systems as well, as Table 1 shows.

TABLE 1: PRINCIPAL CONCEPTS COVERED IN EXISTING HEALTH CLASSIFICATION SYSTEMS¹⁶⁹

<i>Concept</i>	DDR	EQL	15D	HUI ₁	HUI ₂	HUI ₃	IHQOL	HP2000	QWB
Health Perceptions			X					X	
Social Function		X		X			X		X
Speech/Communication			X		X				
Mental Function (Cognitive, Emotional, and/or Affective)		X	X	X	X	X	X		
Physical Function (Mobility, Physical Activity, and/or Self-Care)	X	X	X	X	X	X	X	X	X
Sensory Function or Other Impairments	X	X	X	X	X	X	X		X

Still, there are limits to the inclusiveness of QALYs. The QALY health-classification systems do not define health nearly as inclusively as the World Health Organization (WHO), which has developed a conception of the “quality of life” subsuming twenty-four attributes grouped into six domains—physical, psychological, “independence,” social, environmental, and spiritual.¹⁷⁰ The first three domains cover the territory of QALY health-classification systems, but also include a self-esteem attribute, a body-image attribute, and an attribute for positive as well as negative feelings—all attributes that the QALY systems, and traditional QALY research, treat as background characteristics. The same is true

169. DDR = Disability Distress Ratio; EQL = EuroQOL; 15D = Fifteen Dimension Scale; HUI = Health Utility Index Mark I, II, or III; IHQOL = Index of Health-Related Quality of Life; HP2000 = Years of Healthy Life; QWB = Quality of Well-Being Scale. This is based on a table in Donald L. Patrick & Pennifer Erickson, *Applications of Health Status Assessment to Health Policy*, in *QUALITY OF LIFE AND PHARMACOECONOMICS*, *supra* note 1, at 721. See also *COST-EFFECTIVENESS IN HEALTH AND MEDICINE*, *supra* note 12, at 95 (listing “core concepts and domains of health-related quality of life,” including health perceptions, social function, psychological function, physical function, and impairment).

170. On the development of the WHOQOL, see The WHOQOL Group, *supra* note 90; and Adler, *Fear Assessment*, *supra* note 54, at 1051 n.197.

of the attributes covered by the last three WHO domains. The “social” domain asks about the quality of the subject’s personal relationships, social support, and sex life. The “environment” domain covers personal security, housing quality, wealth, access to information and education, access to social services, recreational activities, pollution, and transport. The spiritual domain asks about the perceived meaningfulness of the subject’s life. Traditional QALY research does not see these domains as constitutive of “health.”

In fact, there is very little that seems relevant to well-being which is not covered by one of the WHO quality of life domains. “Quality of life” ends up being a synonym for “welfare,” and not merely “health,” normally understood as a proper subset of the welfare-affecting features of life-histories. But why not scrap or redefine the concept of “health,” for QALY purposes, so that the set of background characteristics shrinks to zero? Respondents would be asked to use the SG or TTO formula to place states on a 0-1 scale, where 0 is death and 1 is an ideal state not just with respect to physical, emotional, cognitive, and hedonic attributes, but also with respect to body image, self-esteem, personal relationships, social support, sex life, personal security, housing quality, and everything else that makes life worth living.

There are some apparent difficulties with making QALYs this inclusive. Many if not all respondents would presumably find it cognitively overwhelming to make cardinal comparisons (using TTOs or SGs) among states that vary with respect to the totality of welfare-relevant attributes. Current QALY surveys hold fixed most attributes, the non-health ones, and give respondents the more manageable task of making cardinal comparisons of states that vary only with respect to health attributes. Relatedly, QALY valuations elicited from many (if not all) respondents relative to a maximally inclusive QALY scale would be less sensitive to small changes than traditional QALYs.¹⁷¹ States that are not radically different with respect to the physical, cognitive, emotional, and hedonic attributes subsumed by traditional QALY measures, and do not differ at all with respect to other attributes, might be lumped together and assigned the same number on a maximally inclusive QALY scale, but differentiated by traditional QALYs. For example, a recent synthesis of traditional QALY research suggests that 0.75 is a plausible QALY score for angina, and 0.81 for pancreatitis.¹⁷² If 0 is death and 1

171. See generally Brazier et al., *supra* note 1, at 13 (discussing desirable properties of health valuation instruments, including “responsiveness,” which is defined as “the ability of an instrument to measure clinically significant changes in health”). See also Cam Donaldson et al., *Should QALYs Be Programme-Specific?*, 7 HEALTH ECON. 239 (1988) (suggesting that program-specific health scales may be more sensitive than general scales).

172. Cf. Harvard Center for Risk Analysis, Cost-Effectiveness Analysis Registry, <http://www.hpsh.harvard.edu/cearegistry> (database of cost-utility analyses, including 0.75 as one QALY valuation for angina and 0.808 as the valuation of recurrent pancreatitis).

is a state that is ideal with respect to all the dimensions of welfare, would the typical respondent using the TTO or SG method be able to differentiate numerically between an ideal welfare state marred only by angina and an ideal welfare state marred only by pancreatitis, assigning the first (say) a score of 0.92 and the latter a score of 0.93?¹⁷³ It is quite plausible that he would not.

Given these measurement problems, the optimally inclusive QALY scale for purposes of standard surveys eventuating in policy evaluation might well be less than fully inclusive, and exclude some of the attributes that the WHO conception, philosophers of welfare, and ordinary intuition see as relevant to welfare. In any event, existing QALY surveys are much less than fully inclusive, and policy evaluation using these valuations must be sensitive to the fact that they are less than perfect proxies for LWUs, given background characteristics.

D. Non-Health Changes

The final difficulty with the QALY scale is intimately related to the one just described. If some determinants of welfare fall outside the domain of “health,” then some welfare changes will not be measurable using QALYs (equivalently, the measure of the change will be zero). Changes in consumption provide the most practically compelling example. These sorts of welfare effects are the heartland of standard cost-benefit analysis, and are pervasively counted as costs or benefits in actual policy-evaluation practice, but do not register, as such, on the traditional QALY scale.¹⁷⁴ Extravagant consumption plus angina has the same QALY score, *ceteris paribus*, as moderate consumption plus angina—given a QALY scale that counts consumption and wealth as background attributes. A different example comes from environmental regulation. Environmental economists and, increasingly, agencies such as the EPA incorporate the following sorts of nonmarket benefits into cost-benefit studies: the enjoyment experienced by visitors to parks or other protected areas, the recreational benefits of hunting and fishing, the improved visibility that accompanies better air quality, smell- or noise-avoidance, the “scenic” benefit of viewing a nice landscape, and the sheer satisfaction of knowing that a site, ecosystem, or species exists.¹⁷⁵ All these effects will show up in WTP/WTAs amounts, as elicited using contingent-valuation or revealed preference techniques, but none fall within the domain of

173. This calculation assumes that health is one-third of overall welfare and rounds valuations to two digits. If health is less important to overall welfare, the problem is exacerbated.

174. See Hammit, *How Much Is a QALY Worth?*, *supra* note 96, at 5; Schulphar & O’Brien, *supra* note 144, at 214. *But see* COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12, at 122 (recommending that financial consequences related to health status be measured on a QALY scale and be included in the denominator, rather than the numerator, of cost-effectiveness ratios).

175. See Adler, *Fear Assessment*, *supra* note 54, at 981-82.

health, at least as defined by the traditional QALY health-classification systems.¹⁷⁶

V. QALYs, NONTRADITIONAL COST-BENEFIT ANALYSIS, AND WELFARIST POLICY EVALUATION

What role should QALYs play in welfarist policy evaluation, given both their advantages relative to a WTP/WTa metric and their limitations? This Part answers the question in two different ways. First, I sketch a theoretically appealing, but currently unrealistic, decisional approach where LWUs themselves serve as the policy metric, and QALY surveys, WTP/WTa surveys, revealed preference data, and other information is used by the decision-maker in a Bayesian fashion to “update” her estimate of LWUs. Second, I describe an approach that is less elegant but more realistic, given current practices: a nontraditional kind of cost-benefit analysis that incorporates dollar amounts derived from QALY measurements, not merely WTP/WTa amounts, to measure health effects. The FDA has, over the last half-decade or so, followed precisely this approach in some of its important rulemakings—converting QALYs to dollars using a conversion factor ranging from \$100,000 to \$500,000, and then plugging those dollar amounts into a monetized cost-benefit analysis.

The FDA, here, is traversing the frontiers of policy science, and in a way that (I believe) constitutes a true advance. This Part will argue that the FDA’s nontraditional cost-benefit procedure can improve on traditional policy analysis, at least in some choice situations, and will analyze, in a preliminary way, how the procedure is optimally structured. At what rate should QALYs be converted to dollars? When should QALYs be substituted for traditional WTP/WTa measurements of health effects? I conclude by discussing a concrete case where the applicability of traditional cost-benefit analysis has been hotly contested and where QALY-based cost-benefit analysis may well be an improvement: the use of WTP/WTa amounts to measure the cost of premature death.

A. *First-Best Policy Evaluation: LWU Maximization with Some Help from Bayes*

I have argued, to this point, that overall welfare is relevant to policy evaluation; that overall welfare is best understood in terms of the convergent preferences of idealized spectators contemplating extended lotteries; and that LWUs, a numerical scale of these preferences, constitute the welfarist gold-standard. One outcome has greater overall well-being than a second just in case the sum of LWUs is greater in the first case. Correctly measured WTP/WTa amounts are not perfect proxies for LWUs, and in any event the WTP/WTa

176. See Klose, *supra* note 133, at 115 (noting that WTP/WTa amounts, by contrast with QALYs, “provide[] a more comprehensive measure of the effects of a health care technology”).

valuations revealed by behavior or expressed in surveys diverge from true WTP/WTa amounts. Correctly measured QALYs, too, are not generally perfect proxies for LWUs, given the problem of background characteristics; further, the additive formula for estimating QALYs will be inaccurate, if the conditions for additive decomposition break down; and cognitive failures may interfere with QALY surveys, just as they interfere with the measurement of WTP/WTa amounts.

What does this analysis imply for welfarist policy evaluation? It suggests that—placing to one side problems of deliberation costs, including computational limits—the optimal welfarist procedure would use neither dollar amounts, nor QALY amounts, but rather the LWU scale itself to evaluate different policy options. The proposal may seem outlandish—but remember that measurement techniques now intimately familiar to us, such as monetized cost-benefit analysis, are themselves quite new on the scale of human time, invented only one or two generations ago.¹⁷⁷ The science of measurement evolves, and my notion of an LWU scale is in fact well-grounded in current measurement theory, representing the application of utility theory—a theory very widely accepted by contemporary welfare economists—to a particular set of preferences, the spectators’ preferences over life-histories.

To be sure, the spectators are idealized, and this raises a large problem of uncertainty. Who knows what life-histories, and lotteries over life-histories, we would favor, if we were fully informed, fully deliberative about the different life-histories and lotteries being considered, and unhampered by errors in cognition or judgment? Humans’ actual preferences for life-histories and lotteries over life-histories, as evidenced by their verbal or physical behavior, are only very imperfect evidence of their idealized preferences. But policy-analytic techniques for handling uncertainty continue to develop apace, just as measurement techniques do.¹⁷⁸ The best, general approach to uncertainty is the “Bayesian” approach. Given some item of interest whose numerical value is uncertain, the policymaker starts with a “prior” probability distribution over possible numerical values of the item, and then updates her probability distribution as new information arrives.¹⁷⁹ Practicable, statistical techniques for implementing the

177. See Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 167-76 (summarizing the history of cost-benefit analysis).

178. See generally M. GRANGER MORGAN & MAX HENRION, *UNCERTAINTY: A GUIDE TO DEALING WITH UNCERTAINTY IN QUANTITATIVE RISK AND POLICY ANALYSIS* (1990).

179. On Bayesian approaches to probability generally, see Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. 1293, 1312 n.73 (2003) (citing sources). On Bayesian statistics, see, e.g., JOSE M. BERNARDO & ADRIAN F.M. SMITH, *BAYESIAN THEORY* (1994).

Bayesian idea are now available,¹⁸⁰ and these are increasingly employed by policy analysts. For example, such techniques are employed to determine the optimal design of nuclear reactors or other large, dangerous structures in the teeth of uncertainty about the external stresses that the structures will be subjected to and the processes that will occur inside them;¹⁸¹ or to predict the number of deaths that will result from pollution or other environmental releases of toxic chemicals, despite fairly deep uncertainty about dose-response relationships, the environmental “fate and transport” of toxins, and demographic patterns over time.¹⁸²

I suggest that Bayesian techniques could, in principle, be used to estimate LWUs. For a given type of welfare impact (a risk of death, a headache, a scenic view, angina, and so on), a prior probability distribution over the LWU values of that impact would be defined. That distribution might be quite “diffuse”—the analyst might have a very poor initial sense of how idealized spectators would value the impact—or her priors might be less diffuse, incorporating for example her own intuitions about welfare, or philosophical wisdom, or her knowledge about the psychological processes of preference formation (all supporting probabilistic guesses about what the spectators would want).¹⁸³ QALY surveys, contingent-valuation surveys, revealed-preference studies, as well as LWU surveys¹⁸⁴ focused directly on measuring the LWU value of the impact could then

180. See, e.g., ANDREW GELMAN ET AL., *BAYESIAN DATA ANALYSIS* (2d ed. 2004).

181. See, e.g., ROGER M. COOKE, *EXPERTS IN UNCERTAINTY: OPINION AND SUBJECTIVE PROBABILITY IN SCIENCE* 27-41 (1991) (discussing use of Bayesian approaches in analyzing nuclear reactor safety).

182. See Adler, *Against “Individual Risk,”* *supra* note 112, at 1208-10 (discussing Bayesian approaches to risk assessment).

183. See, e.g., GELMAN ET AL., *supra* note 180, at 33-72 (discussing informative and noninformative priors).

184. LWU studies would seek to determine respondents’ utilities for whole life-histories. Unlike QALY surveys, these studies would characterize life-histories in terms of background characteristics as well as health. Unlike contingent-valuation or revealed-preference studies they would seek to measure the respondent’s preferences among different life-histories on a utility scale, not a dollar scale. I see nothing to preclude this sort of LWU study, and indeed it would not be radically different from some existing survey work (for example, the surveys in the QALY literature that ask about preferences over entire health-histories rather than assuming temporal decomposition, see *supra* note 69 and accompanying text; the survey work leading up to WHO’s inclusive quality-of-life index, see *supra* note 170 and accompanying text; and the survey work looking to preferences among different life-saving programs, see *infra* text accompanying notes 238-240). To be sure, as suggested earlier, see *supra* text accompanying notes 171-172, the ordinary respondent might find it more difficult to complete an LWU survey than a standard QALY survey. A successful LWU survey might therefore need to incorporate cognitive aids (for example, visual aids to help the respondents grasp the different life-histories being compared; tutorials in the axioms of rational choice; interviewer interventions to point out inconsistencies in valuations), or

be used to “update” the analyst’s priors—just as, for example, the Bayesian scientist’s prior beliefs about the toxicity of a given chemical compound, grounded in the existing scientific literature, are updated by the results of a particular experiment, such as feeding the chemical to a group of rats and seeing how many die.¹⁸⁵

QALY studies, contingent-valuation studies, revealed-preference data, and LWU studies themselves—“observational” evidence analogous, for the Bayesian, to the rat experiment—will have a substantial effect in changing the analyst’s priors if the individuals who are surveyed, or whose behaviors are examined, are particularly well-informed and deliberative. If the individuals’ deliberational and informational characteristics are less proximate to the spectators’ characteristics, then the studies and data may have a less dramatic impact on the analyst’s probability distribution. In either event the analyst will integrate the sources of information available to her to arrive at a probability distribution over LWUs.

The crucial point, here, is that LWUs need not be known for certain for them to figure in policy choice. Uncertainty about LWUs can, in principle, be expressed probabilistically and updated systematically with observations about what people do or say, and the welfarist rule for policy analysis could be: Pick the alternative with the greatest expected LWUs.

I will not elaborate on the approach just sketched, which asks analysts and decision-makers to use the LWU scale itself as their basic decisional tool, rather than QALYs, WTP/WTAs amounts, or some other proxy measure. First, once deliberation costs and the problems of administrative error and opportunism are taken into consideration, the approach may be welfare-suboptimal¹⁸⁶—although it should be pointed out that increases in computing power have made Bayesian techniques feasible in a much wider range of contexts, and much cheaper, than a generation ago.¹⁸⁷ Second, my primary aim here is to inform current welfarist policy-evaluation practices. Cost-benefit analysis is the dominant technique, at least at federal agencies, and describing a different, nonmonetized, LWU-based technique is not of much use to analysts constrained by current practices, institutional routines, and expectations.

perhaps to be limited to particularly able respondents.

185. See, e.g., Ryan A. Hill, *From Science to Decision-Making: The Applicability of Bayesian Methods to Risk Assessment*, 2 HUM. & ECOLOGICAL RISK ASSESSMENT 636 (1996) (discussing use of new data to update prior beliefs about the carcinogenicity of a chemical).

186. See ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, chs. 3, 4 (arguing that welfare-maximizing procedures are sensitive to deliberation costs and decision-maker error and opportunism); Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 217-18 (same).

187. See, e.g., GELMAN ET AL., *supra* note 180, at xx (noting that increases in computing speed and improvements in computational methods have made Bayesian methods feasible for more complicated models and larger datasets).

B. Second-Best Policy Evaluation: Nontraditional Cost-Benefit Analysis

Current welfarist policy-analytic practices are proxy-based. The best-developed measures of welfare are money and QALYs, neither of which directly represent spectators' preferences over complete life-histories, as an LWU scale would. Further, because money has a much broader range than QALYs, subsuming not just health impacts but all manner of non-health effects,¹⁸⁸ the fact that regulatory agencies generally use a money rather than QALY scale to implement the welfarist mandates of presidential cost-benefit orders and of statutes that require cost-benefit analysis is not surprising.¹⁸⁹ This Section will show how QALYs can play a role within, and improve on, monetized cost-benefit analysis.

Monetizing QALYs and incorporating them into cost-benefit analysis is not a technique much discussed by the scholarly literature on QALYs (where QALYs are generally seen as an outcome measure for cost-effectiveness analysis¹⁹⁰), or by the cost-benefit literature (where Kaldor-Hicks efficiency, not overall welfare, is often taken as the gold standard¹⁹¹). Here, practice outruns scholarship—for, as already mentioned, federal agencies have recently started to convert QALY measures of various health impacts to dollar amounts and to add those sums to WTP/WTAs measures of non-health effects in evaluating policy choices. For short, I will call this approach “hybrid” or nontraditional cost-benefit analysis.

The leader in this area is the FDA, which to date has used hybrid cost-benefit analysis in almost twenty rulemakings.¹⁹² The practice apparently began at the FDA in the early 1990s,¹⁹³ was employed in the huge tobacco rulemaking a

188. See *supra* Section IV.D.

189. See Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 167-76 (describing use of monetized cost-benefit analysis by regulators); *supra* text accompanying notes 8-20 (noting infrequent use of QALYs by regulators, with the exception of the FDA).

190. See *supra* text accompanying notes 25-30.

191. See, e.g., Richard Revesz & Robert N. Stavins, *Environmental Law and Policy*, in THE HANDBOOK OF LAW AND ECONOMICS 5 (A. Mitchell Polinsky & Steven Shavell eds., forthcoming 2006) (“The general view from economics is that other criteria in addition to efficiency [in particular, distributive criteria] can and should be employed by policymakers, but that the existence of such criteria does not invalidate the efficiency criterion, which should remain part of social decision-making.”); sources cited *supra* notes 49-53.

192. See *infra* notes 194-195. This count, to be conservative, excludes the handful of rulemakings in which the FDA has used VSLYs. As mentioned earlier, the EPA has also flirted with VSLYs. See *supra* text accompanying notes 14-15; see also Cass R. Sunstein, *Lives, Life-Years, and Willingness To Pay*, 104 COLUM. L. REV. 205, 252 (2004) (listing regulatory impact statements using life-years or QALYs).

193. In a 1993 rulemaking concerning food standards of identity, the FDA appeared to use the

decade ago,¹⁹⁴ and has accelerated since the late 1990s.¹⁹⁵ To give one illustrative

VSLY approach as one measure of mortality costs. *See* Food Standards: Amendment of the Standards of Identity for Enriched Grain Products To Require Addition of Folic Acid, 58 Fed. Reg. 53,305, 53,310 (Oct. 14, 1993). As discussed below, *see infra* text accompanying notes 233-234, this approach is identical to the use of QALY-to-dollar conversions to value mortality if the VSLYs are not age-adjusted or adjusted for other characteristics such as wealth, and if it is assumed the years lost are perfectly healthy.

194. *See* Regulations Restricting the Sale and Distribution of Cigarettes and Smokeless Tobacco To Protect Children and Adolescents, 61 Fed. Reg. 44,396, 44,576 (Aug. 28, 1996).

195. The FDA has employed QALY-to-dollar conversions to monetize morbidity or mortality for purposes of cost-benefit analysis in the following rulemakings. (Where both final and proposed rulemakings employed conversions, only the final rulemaking is cited.) Performance Standard for Diagnostic X-Ray Systems and Their Major Components, 70 Fed. Reg. 33,998, 34,019-22 (June 10, 2005); Establishment and Maintenance of Records Under the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, 69 Fed. Reg. 71,562, 71,621-22 (Dec. 9, 2004); Prevention of Salmonella Enteritidis in Shell Eggs, 69 Fed. Reg. 56,824, 56,853-55 (Sept. 22, 2004); Requirements Pertaining to Sampling Services and Private Laboratories Used in Connection with Imported Food, 69 Fed. Reg. 23,460, 23,469 (Apr. 29, 2004); Bar Code Label Requirement for Human Drug Products and Biological Products, 69 Fed. Reg. 9120, 9159-61 (Feb. 26, 2004); Final Rule Declaring Dietary Supplements Containing Ephedrine Alkaloids Adulterated Because They Present an Unreasonable Risk, 69 Fed. Reg. 6788, 6837 (Feb. 11, 2004); Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. 41,434, 41,487-89 (July 11, 2003); Administrative Detention of Food for Human or Animal Consumption, 68 Fed. Reg. 25,242, 25,261 (May 9, 2003); Medical Devices; Patient Examination and Surgeons' Gloves; Test Procedures and Acceptance Criteria, 68 Fed. Reg. 15,404, 15,411-13 (Mar. 31, 2003); Current Good Manufacturing Practice in Manufacturing, Packing, or Holding Dietary Ingredients and Dietary Supplements, 68 Fed. Reg. 12,158, 12,229-30 (Mar. 13, 2003); Prior Notice of Imported Food, 68 Fed. Reg. 5428, 5455 (Feb. 3, 2003); Registration of Food Facilities, 68 Fed. Reg. 5378, 5410 (Feb. 3, 2003); Marking Requirements for and Prohibitions on the Reimportation of Imported Food Products That Have Been Refused Admission to the United States, 66 Fed. Reg. 6502, 6508 (Jan. 22, 2001); Hazard Analysis and Critical Control Point (HAACP); Procedures for the Safe and Sanitary Processing and Importing of Juice, 66 Fed. Reg. 6138, 6179-84 (Jan. 19, 2001); Current Good Manufacturing Practice for Blood and Blood Components, 65 Fed. Reg. 69,378, 69,398 (Nov. 16, 2000); Surgeon's and Patient Examination Gloves; Reclassification, 64 Fed. Reg. 41,710, 41,732-36 (July 30, 1999); Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rule To Require Refrigeration of Shell Eggs at Retail and Safe Handling Labels, 64 Fed. Reg. 36,516, 36,522-24 (July 6, 1999); Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. 24,254, 24,258-61 (May 1, 1998); Quality Mammography Standards, 62 Fed. Reg. 55,852, 55,963 (Oct. 28, 1997). In addition, the FDA in a few rulemakings has employed VSLYs to monetize mortality without using the term "QALY." *See* Labeling Requirements for Systemic Antibacterial Drug Products Intended for Human Use, 68 Fed. Reg. 6062, 6076 (Feb. 6,

example: The FDA in 2003 promulgated a regulation requiring food nutrition labels to disclose information about trans fats.¹⁹⁶ The rule was economically “significant” for purposes of Executive Order 12,866, the current Presidential cost-benefit order,¹⁹⁷ and the FDA therefore published and sent to OMB a lengthy cost-benefit document, including both monetary estimates of the rule’s costs (the costs of re-labeling foods, testing them to determine trans fats levels, and reformulating them) and monetary estimates of the rule’s benefits (avoided cases of fatal and nonfatal coronary heart disease).

The FDA monetized these benefits in two different ways, once with QALYs and once without, both showing the rule to have greater monetized benefits than costs.¹⁹⁸ In its QALY-based analysis, the FDA calculated the cost of each nonfatal case as the sum of medical costs, functional disability costs, and pain-and-suffering costs. A QALY value for functional disability and pain-and-suffering was estimated, then converted to dollars using a conversion factor of \$100,000 per QALY. As the agency explained:

[A] recent study . . . estimated . . . that the quality adjusted life year for a [coronary heart disease] survivor was 0.71, which indicates that the annual loss to the victim is 0.29 quality adjusted years. This loss represents the combined effects of functional disability and pain and suffering. FDA assumed that the loss lasts for 13 years, or 8.4 discounted years.¹⁹⁹

The agency concluded that the monetized pain-and-suffering and functional disability cost of each nonfatal case was $0.29 \times 8.4 \times \$100,000 = \$243,600$.²⁰⁰ This amount, plus the per-case medical costs, multiplied by the annual number of nonfatal cases prevented by the rule, was the rule’s annual morbidity-reduction

2003); Iron-Containing Supplements and Drugs, 62 Fed. Reg. 2218, 2243 (Jan. 15, 1997); sources cited *supra* note 193; *infra* text accompanying 233-234 (discussing equivalence of QALY and VSLY approaches under certain conditions).

196. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. 41,434.

197. *See* Exec. Order No. 12,866 §§ 3(f), 6(a)(3), 3 C.F.R. 638 (1993), *reprinted in* 5 U.S.C. § 601 (2000); *see also* Unfunded Mandates Reform Act of 1995, 2 U.S.C. § 1532 (2000) (requiring the preparation of cost-benefit analyses for rules resulting in annual expenditures of \$100 million or more).

198. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,487-90. In the proposed rulemaking, the FDA relied solely on QALY-to-dollar conversions to quantify the benefit. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling, Nutrient Content Claims, and Health Claims, 64 Fed. Reg. 62,746, 62,771-75 (proposed Nov. 17, 1999).

199. Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,488.

200. *See id.*

benefit. The FDA's QALY-based approach to monetizing the mortality-reduction benefit of the rule was to estimate the discounted average years of life lost in each fatal case of chronic heart disease (eight or eleven years, depending on the discount rate) and then multiply this number by the \$100,000 QALY-to-dollar conversion factor, to arrive at a monetized benefit figure per fatal case prevented. That figure, multiplied by the annual number of fatal cases prevented by the rule, yielded the annual lifesaving benefit.²⁰¹

The FDA practice underscores that hybrid cost-benefit analysis is a genuine policy-analytic option. The practice puts this novel decision-procedure on the welfarist's table, as it were. But how should the procedure be structured, from the point of view of overall welfare? There are two fundamental issues here. First, what conversion rate should be used to translate QALYs into dollars? The FDA in the trans fats rule used \$100,000 per QALY as its conversion factor, and also considered what total benefits would be at a conversion rate of \$300,000 and \$500,000 per QALY.²⁰² Elsewhere the agency has used a conversion rate of \$373,000 per QALY,²⁰³ and in valuing short-term morbidity it has repeatedly used a "Quality Adjusted Life Day" value of \$630,²⁰⁴ which implies a conversion factor of \$230,000 per QALY.²⁰⁵ Which of these conversion factors should an agency such as the FDA employ? More fundamentally, what principles should

201. *See id.*

202. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,487-90. For other recent rulemakings in which the FDA has used the trio of \$100/\$300/\$500,000 per QALY to determine possible benefits, see Prevention of Salmonella Enteritidis in Shell Eggs During Production, 69 Fed. Reg. at 56,855; and Final Rule Declaring Dietary Supplements Containing Ephedrine Alkaloids Adulterated Because They Present an Unreasonable Risk, 69 Fed. Reg. at 6842.

203. *See* Medical Devices; Patient Examination and Surgeons' Gloves; Test Procedures and Acceptance Criteria, 68 Fed. Reg. at 15,411; Surgeon's and Patient Examination Gloves; Reclassification, 64 Fed. Reg. at 41,734.

204. *See* Requirements Pertaining to Sampling Services and Private Laboratories Used in Connection with Imported Food, 69 Fed. Reg. at 23,469; Current Good Manufacturing Practice in Manufacturing, Packing, or Holding Dietary Ingredients and Dietary Supplements, 68 Fed. Reg. at 12,230; Marking Requirements for and Prohibitions on the Reimportation of Imported Food Products That Have Been Refused Admission to the United States, 66 Fed. Reg. at 6508; Hazard Analysis and Critical Control Point (HAACP); Procedures for the Safe and Sanitary Processing and Importing of Juice, 66 Fed. Reg. at 6180; Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rule To Require Refrigeration of Shell Eggs at Retail and Safe Handling Labels, 64 Fed. Reg. at 36,523; Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. at 24,261.

205. *See* Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. at 24,261.

guide the choice of factor? Second, what function should the converted amounts play in cost-benefit analysis? Specifically, under what conditions should they displace WTP/WTa amounts, where these are available?

1. *How Should QALYs Be Converted to Dollars?*

The FDA derives a QALY-to-dollar conversion factor from VSL (the “value of statistical life”), a number based on WTP/WTa to avoid the risk of death. For those not familiar with traditional cost-benefit analysis of mortality impacts, the following very quick summary might be helpful. Imagine that subjects in a contingent-valuation or revealed-preference study assign a WTP/WTa amount (for example, \$40) to some small risk of death (for example, a 1-in-100,000 risk). That WTP/WTa amount, divided by the risk, is the “VSL” implied by the study (in this instance \$4 million). It is the cumulative amount that a large population of individuals with the subjects’ preferences would be willing to pay so as to avoid, or willing to accept as compensation for, a single death that would occur for certain, but with an uncertain victim.

The standard cost-benefit technique that agencies currently employ to monetize the mortality effect of a policy is to estimate the total lives saved or lost and multiply that number times a VSL figure inferred from a large group of contingent-valuation and/or revealed preference studies of WTP/WTa for the risk of death.²⁰⁶ The VSL figures actually employed by agencies are in the vicinity of \$6 million.

Now for the question of QALY-to-dollar conversion factors. Imagine that we calculate the average VSL from a group of contingent-valuation or revealed-preference studies. The average life expectancy of the individuals in the studies is X years; life expectancy is discounted at some rate so that the average discounted life expectancy is Y ; and the average health of the group is at QALY level $q < 1$. VSL/qY seems, intuitively, like an appropriate QALY-to-dollars conversion factor—and in any event this has been the approach generally articulated by the FDA,²⁰⁷ as well as the approach taken in scholarly work by some health

206. For example, the agency might determine the VSL implied by each study (average WTP/WTa of the subjects divided by the risk involved), and then average those study-specific VSLs. On the VSL method and the use of VSL by administrative agencies, see Adler, *Against “Individual Risk,”* *supra* note 112, at 1198 n.300.

207. See, e.g., Prevention of Salmonella Enteritidis in Shell Eggs During Production, 69 Fed. Reg. at 56,855; Bar Code Label Requirement for Human Drug Products and Biological Products, 69 Fed. Reg. at 9160; Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,489; Medical Devices; Patient Examination and Surgeons’ Gloves; Test Procedures and Acceptance Criteria, 68 Fed. Reg. at 15,411; Current Good Manufacturing Practice in Manufacturing, Packing, or Holding Dietary Ingredients and Dietary Supplements, 68 Fed. Reg. at 12,230; Current Good Manufacturing

economists.²⁰⁸ VSL/qY is the average WTP/WTa measure of the loss of longevity that the subjects in these studies were at risk of losing, divided by the average (discounted) QALY measure of that longevity.

But intuitions can mislead. The approach I will propose for deriving a QALY-to-dollar conversion factor is, at least conceptually, quite different from the FDA's. My approach flows directly from my view of cost-benefit analysis as a decision procedure maximizing LWUs.

To begin, it makes little sense, I suggest, to think of decision-makers setting a QALY-to-dollar conversion factor on a one-off basis, for some particular choice situation. To identify the factor that maximizes LWUs, the decision-maker needs to have some sense of how QALYs and WTP/WTa amounts translate into LWUs. If she can do *that* on a case-by-case basis with relative ease and accuracy, why not simply analyze the choices at hand directly in terms of LWUs? Why do a cost-benefit analysis of each option, which translates its various welfare effects into money amounts and then aggregates, rather than an LWU analysis, which translates those effects into LWU amounts and then aggregates?

Rather, the choice of the QALY-to-dollar conversion factor is best understood as a problem at the level of systems design. An agency head or an oversight body (the "system-designer") is anticipating that a conversion factor will or might be used in a range of choice situations confronted by agency decision-makers, and is determining what the optimal factor would be. The decision-makers themselves will evaluate choices by performing cost-benefit analysis, not LWU analysis. As noted above, there are various plausible reasons, having to do with deliberation costs and with the competence and trustworthiness

Practice for Blood and Blood Components, 65 Fed. Reg. at 69,368; Surgeon's and Patient Examination Gloves; Reclassification, 64 Fed. Reg. at 41,734; Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. at 24,261. The \$100,000 per QALY conversion factor that the FDA periodically uses is based on research by Garber and Phelps, who employ a different approach to estimating the factor. *See, e.g.*, Prevention of Salmonella Enteritidis in Shell Eggs During Production, 69 Fed. Reg. at 56,855; Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,489 (citing Garber & Phelps, *supra* note 28). A less refined version of the approach described in the text would assume that q is 1. A more refined version would allow q to vary over time. A related approach would use average population longevity rather than the average life expectancy of individuals in WTP/WTa studies for the risk of death. All these variations of the VSL/qY approach are quite different from the method I advocate, in the text below, and are vulnerable to the criticisms of the VSL/qY approach articulated below.

208. *See, e.g.*, Richard A. Hirth et al., *Willingness To Pay for a Quality-Adjusted Life Year: In Search of a Standard*, 20 MED. DECISION MAKING 332, 335 (2000); George Tolley et al., *State-of-the-Art Health Values*, in VALUING HEALTH FOR POLICY, *supra* note 1, at 328-29.

of the decision-makers,²⁰⁹ why it might be LWU-maximizing for the decision-makers to employ cost-benefit analysis rather than LWU analysis as their decision-procedure, at least at present.

However, the system-designer *will* employ LWUs in setting the QALY-to-dollar conversion factor. At a minimum, she will need to have some very rough sense of the expected LWU gains and losses associated with different conversion factors. In the set of choice situations that the system-designer is considering, health effects will be measured by the decision-maker in QALYs and converted to dollar amounts through some conversion factor f , while non-health effects will be measured on a WTP/WTa scale.

For simplicity, think of each choice situation as binary, presenting the decision-maker with a choice between the status quo and some regulatory intervention or “project,” which has health benefits and non-health costs. Ideally, for any choice situation in the set, the system-designer would be able to express probabilistically her judgments about the total amount of the project’s health benefits and non-health costs, in LWUs, plus her judgments about the likely ratio between the total WTP and QALY amounts that the agency decision-maker will observe and the total LWU amounts.²¹⁰ For simplicity, I will also assume that the system-designer’s subjective probability distributions over project LWU benefits and costs, over the QALY/LWU ratio, and over the WTP/LWU ratio are the same for every choice situation in the set, represented by the following sorts of

209. *See supra* text accompanying note 186.

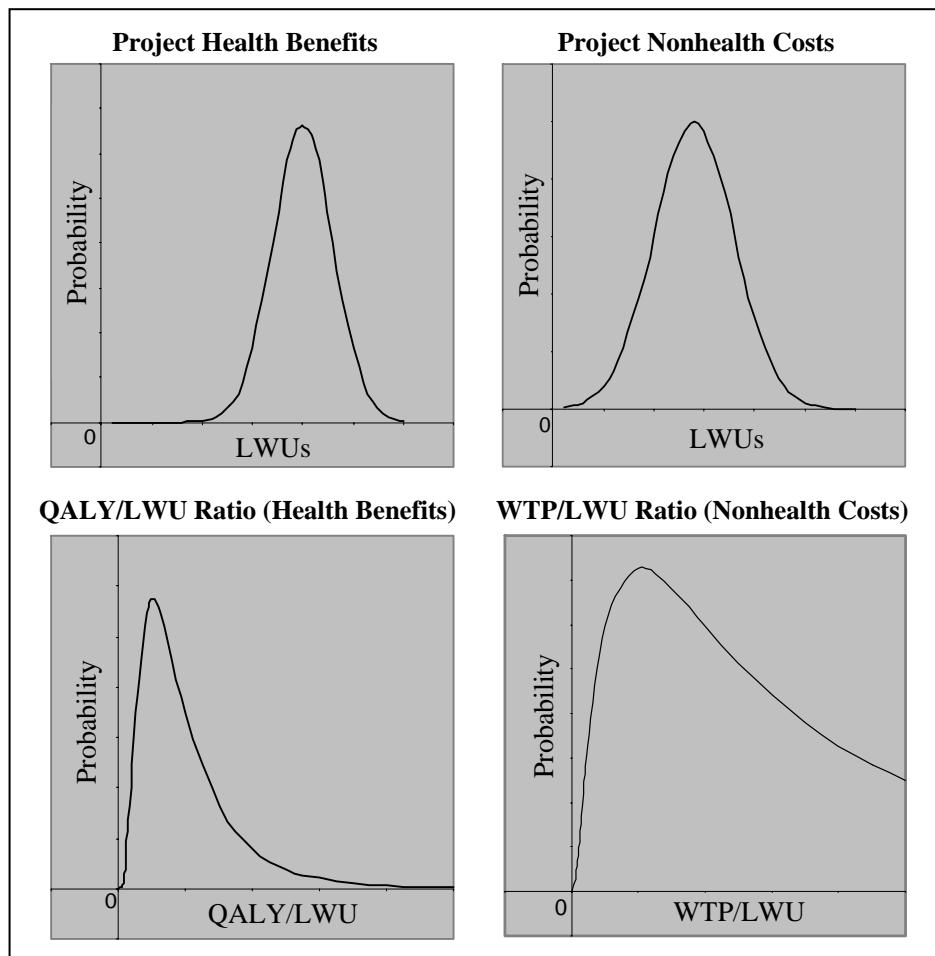
210. To be clear: The WTP/LWU and QALY/LWU ratios here—the ratios that drive the system-designer’s choice—are the ratios between the aggregate WTP or QALY amounts for project costs or benefits that the decision-maker will calculate, and the total amount in LWUs of project costs or benefits. These are ratios of total project impacts, not the average of the QALY/LWU and WTP/LWU ratios for the various individuals affected by the project (although under some conditions the average of individual ratios may equal or approximate the ratio of totals).

The graphs in the text are easiest to grasp if it is assumed that the QALY/LWU ratio for health benefits is independent of the amount of health benefits in LWUs, and similarly that the WTP/LWU ratio for non-health costs is independent of the amount of non-health costs in LWUs. In that event, the system-designer has a single subjective probability distribution for the QALY/LWU ratio and WTP/LWU ratio conditional on any given level of benefits and costs (in LWUs), and the ratio graphs display this distribution. However, the system-designer’s maximization problem as modeled here does not require this sort of independence. *See infra* App. II.

The analysis does make a different sort of independence assumption, namely that the system-designer’s probability distributions with respect to LWUs, the QALY/LWU ratio, and the WTP/LWU ratio for a given choice situation remain the same regardless of which choice the decision-maker selects in any other choice situation. For example, the designer’s probability distributions with respect to the second choice that the decision-maker will confront are the same regardless of whether the decision-maker has picked the project or status quo in the first-choice situation.

graphs. A more comprehensive analysis would permit the designer to have different subjective distributions for different subsets of choice situations; but that analysis would be overly complicated for my purposes here, which is to explain a general approach to setting the QALY-to-dollar conversion factor.

FIGURE 1: SETTING A QALY-TO-DOLLAR CONVERSION FACTOR



Note that both the QALY/LWU ratio for health benefits, and the WTP/LWU ratio for non-health costs, are variable. This reflects the fact that the actual QALY or WTP amounts measured by the decision-maker will not be perfect proxies for LWUs—for all the reasons already discussed. A perfect proxy would have some single fixed ratio to LWUs with probability one.

With these probability distributions in hand, the system-designer can pick an

optimal QALY-to-dollar conversion factor f^* . This is the factor f^* that maximizes expected LWUs. To reiterate: The system-designer anticipates that decision-makers will perform hybrid cost-benefit analysis across a range of choice situations, with health effects measured in QALYs and converted to dollars using a conversion factor that the system-designer will announce, and non-health effects measured in WTP/WTAs. Her underlying objective is to maximize overall (expected) welfare,²¹¹ i.e., total expected LWUs. The optimal conversion factor f^* is the factor that does just that. Cost-benefit analysis, as I see it, is simply a tool for advancing overall well-being, and should be calibrated (e.g., in the choice of the QALY-to-dollar conversion factor) with that objective in mind.

What can we say about f^* , the optimal QALY-to-dollar conversion factor? An Appendix to this Article derives some formal results about f^* , which are informally summarized here. First, f^* is context-dependent. It depends on the shapes of all the probability distributions displayed here: on the system-designer's probability distributions with respect to the QALY/LWU ratios for health benefits and WTP/LWU ratios for non-health costs, as well as her probability distributions with respect to the total amount of project health benefits in LWUs, and total project non-health costs in LWUs.²¹² This may seem like a negative result, but in fact it underscores the crucial point that specifying the QALY-to-dollar conversion factor is a pragmatic decision, a matter of optimizing hybrid cost-benefit analysis with respect to the underlying criterion of expected overall welfare, representable in expected LWUs. The optimal factor for one set of choice situations may be f^+ , given the system-designer's QALY/LWU, WTP/LWU, and LWU probability distributions for that set of situations. The optimal factor for a different set of choice situations, e.g., a different administrative agency, (or for the same set of choice situations but a different designer, with different probability distributions) may be f^{++} . There is no single, natural, acontextual "rate" at which QALYs convert into dollars.²¹³

Second, one can show that, with some simplifying assumptions, the optimal conversion factor in a given context, for a given set of QALY/LWU, WTP/LWU and LWU probability distributions, is a kind of *average*, which can be calculated by taking the average WTP/LWU ratio for non-health costs, and dividing that by

211. The system-designer does not know for certain what choices will be presented to the decision-makers, and what the WTP/LWU and QALY/LWU ratios will be. Proceeding, therefore, under uncertainty, she aims to maximize expected LWUs.

212. See *infra* App. II.

213. In a recent article, Gyrd-Hansen proposes a "pragmatic" approach to determining WTP per QALY, which is similar in spirit to my proposal. He suggests that "seeking to apply a unique WTP for a QALY should not be seen as defining the theoretical link between CEA and CBA, but rather as an aid to decisionmakers," and notes that decision-makers might use "situation-specific" WTP per QALY values. Dorte Gyrd-Hansen, *Willingness To Pay for a QALY: Theoretical and Methodological Issues*, 23 PHARMACOECONOMICS 423, 428, 430 (2005).

the average QALY/LWU ratio for health benefits.²¹⁴

Conceptually, this approach to setting the QALY-to-conversion factor is quite different from the FDA's *VSL/qY* approach. My approach seeks to calibrate the conversion factor in terms of LWUs; the FDA's does nothing of the sort, at least not explicitly.

Now, it might be objected that FDA's approach does implicitly incorporate LWUs. Leaving aside discounting, the FDA in effect takes *VSL*, which is the WTP/WTa measure of a certain kind of health impact, namely death, and divides that by a kind of estimate of the average QALY loss that occurs in death, namely *qX*. But this WTP/QALY ratio, VSL/qX , mathematically, is identical to: $(VSL/L)/(qX/L)$, where *L* is the LWU loss that on average occurs with death. In other words, the FDA's conversion factor can be seen as a rough kind of approximation for the number that would emerge if a system-designer calculated the average WTP/LWU ratio for the mortality-reduction benefits of policy choices, and divided that by the average QALY/LWU ratio for those choices.

Even so, there is a crucial difference between the FDA's approach and mine. My analysis shows that, with simplifying assumptions, the optimal conversion factor is the average WTP/LWU ratio for the *non-health* impacts of agency choices, e.g., reduced consumption, or recreation, or employment, divided by the average QALY/LWU ratio for health impacts. By contrast, the FDA takes the average WTP/LWU ratio for a particular *health* impact (the risk of death) and divides that by the average QALY/LWU ratio for that impact. Think of the point this way: Hybrid cost-benefit analysis converts QALYs into dollars and then adds these dollar sums to WTP/WTa amounts for non-health impacts. So what is crucial in optimizing the conversion is how WTP/WTa amounts for non-health impacts translate into LWUs and how QALYs translate into LWUs, *not* how WTP/WTa amounts for health impacts translate into LWUs.

A final attempt to salvage the FDA's approach: Assume that the distribution of WTP/LWU ratios for health impacts is roughly the same as the distribution of WTP/LWU ratios for non-health impacts. On that assumption, the FDA's approach is perhaps a rough and ready way to approximate the optimal conversion factor—although even here one would want to consider WTP/LWU ratios for a range of health impacts, not simply the risk of death (as the FDA does). Doing that suggests that the range of numbers the FDA has employed (\$100,000 to \$500,000) per QALY is on the high side; a conversion factor of \$100,000 per QALY looks closer to optimal, and lower factors such as \$50,000 or even \$10,000 should be considered.²¹⁵ But in any event the assumption may

214. See *infra* App. II.

215. Hirth et al. examined a wide range of VSL estimates, based on contingent valuation studies, wage-risk studies, and other revealed preference studies. Their estimates of the conversion

not be true, which means that the numbers emerging from this quite rough and ready analysis would need to be adjusted. For example, the biases that affect WTP/WTa measurement may tend to inflate WTP/LWU ratios for health as opposed to non-health impacts, or vice versa. I will not attempt to estimate the degree or direction of the appropriate adjustment here.

It might be objected that my approach requires the system-designer to do something impossible: to estimate project impacts in an esoteric and unobservable metric, namely LWUs. But that endeavor is *not* impossible—although it may be expensive and time-consuming, which is why it may well be welfare-maximizing for deliberation about LWUs to occur at the level of systems design, rather than in the evaluation of individual projects. The graphs in this Section, again, represent the designer's subjective probability distributions—her

factor were in the same range as the FDA's numbers: \$93,402 based on the revealed preference studies other than wage-risk studies, \$161,305 based on the contingent valuation studies, and \$428,286 based on the wage-risk studies, or an overall estimate of \$265,345. *See* Hirth et al., *supra* note 208, at 338. By contrast, the conversion factor that results from comparing QALY values and WTP/WTa amounts for light morbidity is substantially lower. I estimated this factor using a review article by Johnson, which collects WTP/WTa and QALY values for a range of light symptoms, such as angina, throat congestion, coughs, runny noses, and headaches. *See* Johnson et al., *supra* note 129. The estimation procedure was straightforward. If, for example, a study determines that average WTP to avoid seven days of a severe cough and sneeze is \$87.35, and the QALY loss associated with a severe cough and sneeze is 0.318, then the conversion factor implied by this study is $(365 \times 87.35) / (7 \times 0.318)$, or \$14,323. Averaging the conversion factors implied by the fifty-some WTP/WTa and QALY valuations for light morbidity produces an overall estimate of \$37,663 per QALY. This is an order of magnitude lower than the overall Hirth et al. result of \$265,345 per QALY, and (interestingly) much closer to the number traditionally used by many public health scholars as the cut-off ratio for cost-effectiveness analysis: \$50,000 per QALY. *See* Hirth, *supra*, at 333. It is also very close to the cut-off ratio that, in practice, the Australian government uses in deciding whether to list a pharmaceutical for public funding, which one study estimates to be \$40,400. *See* Dorte Gyrd-Hansen et al., *Willingness To Pay for a QALY*, 12 HEALTH ECON. 1049, 1049 (2003).

Gyrd-Hansen and colleagues come up with a yet lower estimate of the conversion factor. They surveyed 3201 Danish individuals and elicited WTP values for changes in health states described using a standard health state classification system. These money valuations were integrated with preexisting QALY valuations for the changes, yielding a mean WTP per QALY of roughly \$10,000. *See id.* at 1058. A simple average of this figure, the \$37,663 ratio implied by the Johnson article, and the \$265,345 overall estimate based on VSL set forth by Hirth et al. gives a value of about \$100,000. This rough, heuristic calculation suggests that \$100,000, currently the low end of the FDA's range, might well be an appropriate central QALY-to-dollar conversion factor for hybrid cost-benefit analysis—absent further information, for example information concerning the distribution of LWU costs and benefits across agency choice situations, that would bear on the optimal conversion factor—and that lower factors, such as \$50,000 or even \$10,000, should be considered.

degrees of belief about the different possible quantities of some item. An individual's subjective probability distributions for *anything* (the temperature in the middle of the sun, the number of electrons in George Washington's finger, the number of mistakes that students taking a hypothetical exam under hypothetical conditions will make) can be generated using Bayesian probability-elicitation techniques, at least if the individual is smart and patient enough to go through the exercise.²¹⁶

LWUs are utility numbers representing the preferences of idealized spectators with respect to the different combinations of health, consumption, and other attributes that make up different possible human life-histories. LWUs are not observable; but neither are QALYs. LWU surveys, asking respondents to think about, and express numerically, their preferences over possible lives, can and certainly should be conducted.²¹⁷ Even without such surveys in hand, system-designers can arrive at a sense of how strongly spectators would prefer different packages of health and non-health attributes by consulting their own preferences, and by drawing on the rich body of economic and philosophical scholarship about well-being.²¹⁸ Sketching a subjective probability distribution over LWUs means sketching a subjective distribution over the utilities of hypothetical, idealized spectators. Those utility numbers—like physical magnitudes in inaccessible places, or the quantities of miniscule items or items in the past or future—cannot be perceived. But we can at least quantify our beliefs about what the numbers might be, in the form of subjective probabilities. Once it is recognized that prevalent policy metrics, such as QALYs and WTP/WTa amounts, are simply rough proxies for well-being, it becomes plausible—indeed compelling—to undertake more systematic efforts to estimate LWUs, and to use such estimates in structuring cost-benefit analysis. That is what I am advocating here.

2. *What Role Should QALYs Play in Nontraditional Cost-Benefit Analysis?*

What precise role should QALYs play in cost-benefit analysis? At a minimum, QALYs can provide alternate estimates of health benefits. Where both a QALY-based money estimate and a traditional WTP/WTa estimate are available, the agency should undertake cost-benefit analyses using both figures.²¹⁹ If the parallel analyses reach convergent policy recommendations, then

216. On Bayesian techniques, see *supra* text accompanying notes 178-185.

217. See *supra* note 184 and accompanying text.

218. See *supra* text accompanying notes 183-185.

219. The FDA sometimes follows this approach, undertaking both traditional and hybrid cost-benefit analysis. See, e.g., Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. 41,434, 41,488. (July 11,

the agency can be especially confident in that course of action. If they don't, then the agency has reason to scrutinize the WTP/WTa numbers, perhaps conducting additional contingent-valuation or revealed-preference studies. In effect, cost-benefit analysis in parallel provides a simple heuristic of the value of further expenditures to estimate WTP/WTa amounts. It is a kind of rough-and-ready value-of-information analysis.²²⁰ The OMB now recommends that agencies conduct a kind of hybrid cost-benefit analysis as a source of alternate cost-benefit estimates, at least where longevity is at issue.²²¹

Second, agencies should use QALY-based money estimates of health benefits where WTP/WTa estimates are not available. This is not a minor point.²²² There is a large empirical literature on WTP/WTa for mortality risks,²²³ but much less research on WTP/WTa for morbidity;²²⁴ and many, perhaps most, health conditions lack even a single contingent-valuation or revealed preference study. By contrast, as evidenced by the Harvard Center for Risk Analysis's comprehensive compilation of QALY estimates,²²⁵ QALY surveys have been conducted for a large number of conditions. Even where no survey exists for a particular health state, health classification systems often permit researchers to extrapolate a valuation for that state from other, surveyed states.²²⁶

What about a more robust role for QALYs? When, if ever, should an agency prefer a QALY-based money estimate to a WTP/WTa amount? The general welfarist strategy for answering this question is the same as the general strategy for identifying an appropriate QALY-to-dollar conversion factor: Hybrid cost-benefit analysis should be preferred to traditional cost-benefit analysis, over some range of choice situations, when that increases expected welfare (in LWUs). Again, the conditions under which this is true will depend on the sorts of choice situations involved: specifically, the system-designer's probability

2003).

220. On value of information analysis, see, e.g., Maxine Dakins, *The Value of the Value of Information*, 5 HUM. & ECOLOGICAL RISK ASSESSMENT 281 (1999). On the usefulness of heuristics, see generally GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999).

221. OFFICE OF MGMT. & BUDGET, *supra* note 18, at 30.

222. A dramatic example is provided by the EPA's cost-benefit analyses in the major arsenic and radon rulemakings, which employed WTP to avoid chronic bronchitis as an estimate of the cost of nonfatal cancers, since no WTP data for nonfatal cancers was available. See National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, 66 Fed. Reg. 6976, 7012 (Jan. 22, 2001); National Primary Drinking Water Regulations; Radon-222, 64 Fed. Reg. 59,246, 59,325 (Nov. 2, 1999).

223. See sources cited *supra* note 56.

224. See Johnson et al., *supra* note 129, at 642 ("[T]he literature providing monetary health values is deficient in both breadth and quality.").

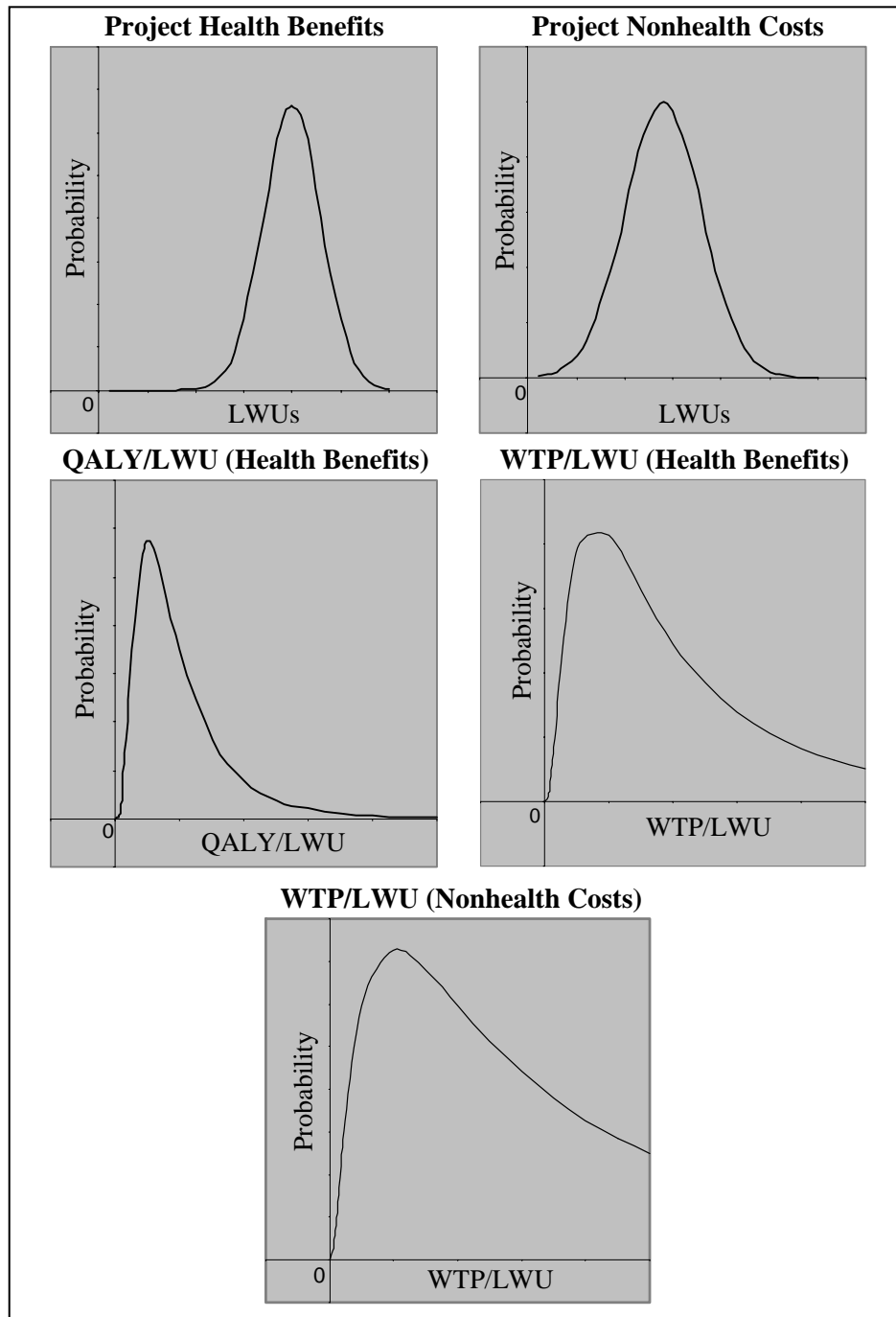
225. See Bell et al., *supra* note 167.

226. See sources cited *supra* note 166 (discussing health state classification systems).

distributions over QALY/LWU ratios, WTP/LWU ratios, and LWU amounts for these choice situations.

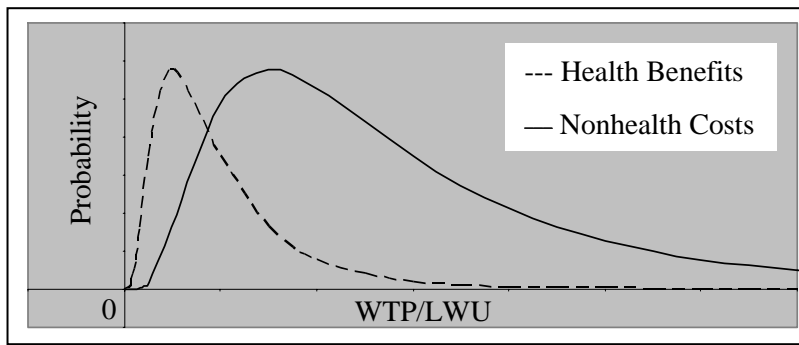
But it is possible to make some general observations. Let us return to the simple scenario discussed above. A system-designer is considering a range of binary choice situations that agency decision-makers will face, each involving a regulatory project with health benefits relative to the status quo and non-health costs. The designer supposes that the non-health costs will be measured using WTP/WTAs and is now choosing between the following two decision procedures: traditional cost-benefit analysis, where the health benefits are measured on a WTP/WTA scale and added to the non-health costs, and hybrid cost-benefit analysis, where the health benefits are measured on a QALY scale, converted to dollars at the optimal conversion rate f^* , and then added to the non-health costs. The designer chooses between these two procedures by determining which one maximizes expected LWUs, and instructs the decision-makers to use that procedure.

FIGURE 2: CHOOSING BETWEEN TRADITIONAL & HYBRID COST-BENEFIT ANALYSIS



There are two general reasons why hybrid cost-benefit analysis might produce higher expected LWUs than traditional cost-benefit analysis. First, the distribution of the WTP/LWU ratio for health benefits might be skewed to the right or left of the optimal point. In other words, a variant of the traditional approach in which health benefits are first measured in WTP/WTAs, then multiplied by a constant k , and finally added to non-health costs, also measured in WTP/WTAs, might lead to better results than straight traditional cost-benefit analysis. This might occur, for example, if the average WTP/LWU ratio for health benefits is greater or less than the average WTP/LWU ratio for non-health costs.

FIGURE 3: WTP/LWU RATIOS FOR PROJECT COSTS & BENEFITS

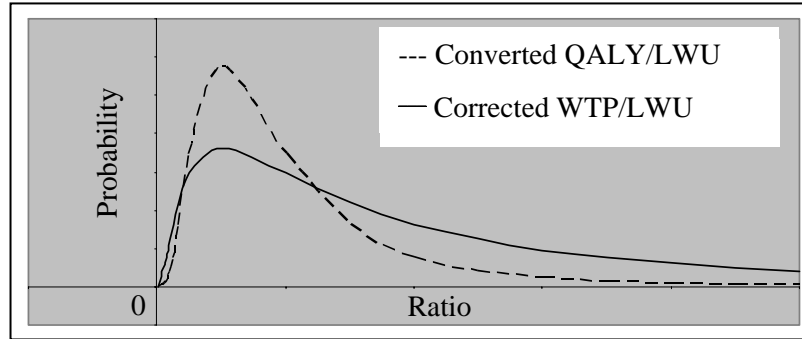


In principle, the skewing of the distribution of WTP/LWU for health benefits, relative to the optimal point, can be corrected by applying a scaling factor to the WTP amounts.²²⁷ Thus, at least in principle, the system-designer should compare (1) a cost-benefit procedure that measures health in QALYs, and converts those amounts to dollars using the optimal conversion factor f^* to (2) a cost-benefit procedure that measures health on a WTP/WTA scale, which is then corrected using the optimal correction factor k^* .

In this competition, the QALY-based metric might win out because it has smaller variance or spread.

227. Cf. Hirth et al., *supra* note 208, at 340 (discussing use of “calibration factors” to correct WTP amounts).

FIGURE 4: CONVERTED QALY/LWU & CORRECTED WTP/LWU RATIOS,
PROJECT HEALTH BENEFITS



Take the simplest case in which all non-health costs measured in WTP convert to LWUs at the same rate and in which the designer's QALY/LWU distribution and WTP/LWU distribution for health benefits are independent of the absolute level of benefits and costs, in LWUs. In this simple case, if the ratio of optimally converted QALYs to LWUs has less “area under the tails” (roughly, less variance) than the ratio of optimally corrected WTPs to LWUs, then the QALY metric is better.²²⁸ The Appendix demonstrates this rigorously. For each pairing of health benefits, in LWUs, and non-health costs, in LWUs, the lower variance of the QALY/LWU distribution means a lower probability that the QALY metric will reach the incorrect result (that is, choosing the project where health benefits are less than costs, or choosing the status quo where health benefits exceed costs).

The idea of variance provides a unifying rubric under which to group the diverse ways in which QALYs and WTP amounts can fail to track LWUs. Both (1) the cognitive difficulties that drive a wedge between the QALY or WTP valuations available to regulators and idealized QALYs or WTP amounts, and (2) the factors that drive a wedge between idealized QALYs or WTP amounts and LWUs (such as the wealth effect, dead-anyway effect, and the problem of background characteristics) can be conceptualized as sources of increased variance in the QALY/LWU and WTP/LWU ratios.²²⁹ This observation also underscores the point that the choice between hybrid cost-benefit analysis and traditional cost-benefit analysis is not clear-cut. It represents an exercise in

228. See *infra* App. II.

229. To be sure, these various factors and cognitive difficulties determine not only the variance of the QALY/LWU and WTP/LWU distributions, but also the location of these distributions along the x-axis, i.e., the absolute level of the average QALY/LWU and WTP/LWU ratio. But that effect, in principle, can be compensated for through the conversion factor f^* and correction factor k^* .

optimization, hinging in part on the variance in QALY/LWU and WTP/LWU ratios in the relevant choice situations. Because the factors producing variance are different for the two sorts of health metrics (QALYs and WTP), there is no reason to think that the QALY/LWU variance will always be smaller or larger than the WTP/LWU variance, and thus no reason to think that hybrid cost-benefit analysis will always dominate traditional cost-benefit analysis, or vice versa.

C. Why Not Cost-Effectiveness Analysis?

I have argued that a policy-analytic procedure which employs a “hybrid” cost-benefit analysis, supplementing or displacing traditional cost-benefit analysis, may increase overall welfare (LWUs) as compared to traditional cost-benefit analysis. The observant reader may wonder whether cost-effectiveness analysis should also be back on the table at this point. As discussed earlier, welfarists often criticize cost-effectiveness analysis because it deviates from traditional cost-benefit analysis and Kaldor-Hicks efficiency. Now that our underlying goal is overall welfare, not Kaldor-Hicks efficiency, might not some version of cost-effectiveness analysis be the optimal procedure for health policy?

The answer is pretty clearly no. Cost-effectiveness analysis, again, is a set of policy-analytic techniques that monetize non-health impacts but not health, instead measuring health on some non-monetary scale such as QALYs. The first variant of cost-effectiveness analysis maximizes health for a given budget. Here, the shift from Kaldor-Hicks efficiency to overall well-being as the underlying goal does nothing to salvage the procedure, because it is quite possible that the net welfare benefits of a health-maximizing exhaustion of the budget are less than the net benefits of a smaller or even null expenditure. Cost-benefit analysis tests for that, while cost-effectiveness analysis does not.

The second variant, recall, uses a fixed cut-off ratio to select the best choice. The problem with this approach, as compared to hybrid cost-benefit analysis, is *not* the use of a single cut-off number. Hybrid cost-benefit analysis also does that, in the form of a single QALY-to-dollar conversion factor. Indeed, hybrid cost-benefit analysis employing a conversion factor f^* and the cut-off ratio variant of cost-effectiveness analysis employing that same number f^* as the cut-off would, I believe, always reach the same results.

The problem with this second variant of cost-effectiveness analysis, rather, is the abandonment of WTP/WTAs valuations of health. I have suggested that WTP/WTAs valuations of health impacts and QALY-to-dollar conversions *both* have a useful role in guiding policy choice. First, as stated above, the fact that traditional cost-benefit analysis and hybrid cost-benefit analysis produce divergent recommendations in some policy situation has informational value: It may point to the need for further studies to produce better estimates of WTP/WTAs or QALY values. Second, given the system-designer’s probability

distributions over WTP/LWU, QALY/LWU, and LWU amounts, traditional cost-benefit analysis using WTP/WTa to value health may end up being the welfare-maximizing procedure in some contexts, as compared to a hybrid procedure that employs QALY-to-dollar conversions to value health.

In short, I suggest that the welfare-maximizing procedure will use QALY-based analysis alongside traditional cost-benefit analysis and, where they conflict, will sometimes (not always) prefer the results of the latter analysis. Cost-effectiveness analysis relies solely on QALYs to value health, but since QALYs are not perfect proxies for LWUs, any more than WTP/WTa amounts, that seems overly rigid from the point of view of overall welfare.

D. An Example: The Valuation of Lifesaving

What version of cost-benefit analysis, traditional or hybrid, should be employed to value regulatory measures that cause deaths or save lives? This question has recently been a matter of considerable political controversy and concomitant scholarly discussion, triggered by the EPA's cost-benefit analysis in connection with the "Clear Skies" legislative initiative. The EPA in an alternative estimate used an age-adjusted VSL figure: specifically, \$3.7 million for deaths of individuals under seventy, and \$2.3 million for deaths of individuals over seventy.²³⁰ Vociferous protests by senior citizen groups ensued,²³¹ prompting the OMB to issue a memorandum instructing the EPA and other agencies not to use age-adjusted VSL.²³² In the same memorandum, the OMB discusses valuation of death or life-saving through "VSLYs," that is, calculating the numbers of life-years lost or saved and multiplying by a conversion factor to yield a dollar amount. If the life-years are priced at a *constant* conversion factor, and if the

230. See U.S. EPA, TECHNICAL ADDENDUM, *supra* note 15, at 35-37. Recall that VSLs are derived from WTP/WTa to avoid small risks of death. If an individual is WTP \$3 to avoid a 1 in 1 million risk of death, then the VSL figure implied by that valuation is \$3 million. Cost-benefit analysis with age-invariant VSLs looks at the average VSL, observed in the entire universe of contingent valuation and revealed preference studies (involving subjects of different ages), and employs that single figure (say, \$6 million) to value each death. The age-adjusted VSL methodology seeks to determine the average VSL of individuals in a particular age group, and then uses that age-specific VSL number to value deaths in that age group. For a good discussion of VSLs, age adjustments, and VSLYs, see Lowenstein & Revesz, *supra* note 15.

231. For a description of the controversy, see Lowenstein & Revesz, *supra* note 15, at 10,957; Robert W. Hahn & Scott Wallsten, *Whose Life Is Worth More? (And Why Is It Horrible To Ask?)*, WASH. POST, June 1, 2003, at B03.

232. Memorandum from John D. Graham, U.S. Office of Mgmt. & Budget, to the President's Mgmt. Council (May 30, 2003) [hereinafter Graham Memorandum], available at http://www.whitehouse.gov/omb/inforeg/pmc_benefit_cost_memo.pdf; accord OFFICE OF MGMT. & BUDGET, *supra* note 18, at 30.

years lived would be years of perfect health, then this “VSLY” method is just the kind of QALY-based or hybrid cost-benefit analysis that I have been discussing.²³³ However, the OMB memorandum encourages agencies to use an age-adjusted VSLY approach (in particular, to use a dollar conversion factor that would price life-years saved or lost at a higher amount for senior citizens than for younger individuals).²³⁴

In short, four different cost-benefit approaches to valuing life are now on the table: (1) using an age-invariant VSL figure (e.g., \$6 million), which is standard practice; (2) using an age-adjusted VSL figure; (3) using QALY-to-dollar conversions, i.e., converting each life-year lost or saved by regulation to dollars at an age-invariant conversion rate; and (4) using age-adjusted VSLYs, i.e., converting life-years to dollars at an age-specific rate. The OMB memorandum in response to the Clear Skies controversy, and more recently its authoritative general guide to agency cost-benefit analysis, discourages QALY-to-dollar conversions as well as age-adjusted VSLs.²³⁵

What is the right approach? If Kaldor-Hicks efficiency is the underlying criterion, then age-adjusted VSLs are ideal, at least if they can be measured accurately and cheaply. A decision procedure that sums fully individualized WTP/WTAs amounts will track potential Pareto improvements.²³⁶ Measurement problems may push in the direction of age-invariant VSLs, which represent an average of VSLs across age groups. But VSLs of some kind are preferable to VSLYs (either age-invariant, i.e., QALYs, or age-adjusted VSLYs), since the use of VSLYs to value mortality has no grounding in Kaldor-Hicks efficiency or the WTP/WTA methodology. In short, conventional economic wisdom prefers a VSL measure and is skeptical of QALY-to-dollar conversions and age-adjusted VSLYs.²³⁷

233. See generally Lowenstein & Revesz, *supra* note 15 (discussing life-years approach); Sunstein, *supra* note 192 (same).

234. Graham Memorandum, *supra* note 232; accord OFFICE OF MGMT. & BUDGET, *supra* note 18, at 30.

235. See sources cited *supra* note 232. By recommending age-adjusted VSLYs, OMB discourages age-invariant VSLYs, i.e., straight QALY-to-dollar conversions.

236. This statement is only roughly true—a point that traditional proponents of cost-benefit analysis often overlook—because of the Boadway paradox. See source cited *supra* note 24.

237. See, e.g., Lowenstein & Revesz, *supra* note 15, at 10,963-69. The VSLY method, at least in its simplest variant, values increments to longevity at a linear rate: either a single rate for the entire population (age-invariant VSLYs), or a rate specific to an age group (age-adjusted VSLYs). See *infra* note 246. In either event, a policy that adds five years of longevity (say) to the lives of some individuals in a particular age bracket will be assigned a monetary value five times that of a policy that adds one year. But WTP/WTAs for the risk of losing five years of life need not be five times WTP/WTAs for the risk of losing one year of life, given wealth effects, horizon effects, the

But what if overall welfare, not Kaldor-Hicks efficiency, is the gold standard? The answer, then, is trickier.

Consider, first, the case in which the system-designer (OMB) is choosing an optimal procedure for valuing life across a range of choice situations in which all the individuals whose lives are at stake fall in roughly the same “background equivalent class”: They have equally valuable non-health attributes. A crucial question for this case (and for the more general analysis too) is whether spectators are risk-prone, risk-neutral, or risk-averse with respect to longevity²³⁸—that is, whether the LWU value of an incremental year increases, stays constant, or decreases as affected individuals become older. It seems intuitively plausible that longevity, like wealth, has a diminishing marginal impact on overall well-being; and this assumption is bolstered by some survey work. Johannesson and Johannsson surveyed 1000 Swedes for their policy preferences as between equally costly programs that differed in the number of lives saved and the age of the persons saved. Integrating the survey responses with life expectancy data, they determined that: “[T]hree life-years gained for 50-year-olds are judged equivalent to one life-year gained for 30-year-olds, and ten life-years gained among 70-year-olds are judged equivalent to one life-year gained for 30-year-olds.”²³⁹ Cropper et al. reached similar results in an earlier U.S. study: “[T]he median respondent in our surveys places more weight on saving young persons than he would if people were weighted strictly by life expectancy.”²⁴⁰

If longevity has a diminishing marginal impact on overall well-being, then the LWU value of increments to longevity is smaller as individuals get older, and is a sublinear function of years saved. The monetized QALY value of increments

dead-anyway effect, and so on. Therein lies the basic criticism of VSLYs by those who see cost-benefit analysis as a tool to implement Kaldor-Hicks efficiency.

238. See *supra* text accompanying notes 156-159.

239. Magnus Johannesson & Per-Olov Johannsson, *Is the Valuation of a QALY Gained Independent of Age? Some Empirical Evidence*, 16 J. HEALTH ECON. 589, 595 (1997).

240. Maureen L. Cropper et al., *Preferences for Life Saving Programs: How the Public Discounts Time and Age*, 8 J. RISK & UNCERTAINTY 243, 244 (1994). Cropper does not find that the utility of lifesaving decreases continuously with the age of the person saved. Rather, “[e]ight 60-year-olds are judged equivalent to saving one 20-year-old” and “eleven 60-year-olds are judged equivalent to saving one 30-year-old,” suggesting “that the utility attached to saving an anonymous life is a hump-shaped function of the age of the person saved.” *Id.* at 244-45. For a parallel survey focused on the social value of health improvements at different ages, see Jan J.V. Busschbach et al., *The Utility of Health at Different Stages in Life: A Quantitative Approach*, 37 SOC. SCI. & MED. 153 (1993). A recent review of studies asking respondents to prioritize health benefits for different members of the population finds that “most studies suggest that health gains to the old are weighted less.” Paul Dolan et al., *QALY Maximisation and People’s Preferences: A Methodological Review of the Literature*, 14 HEALTH ECON. 197, 202 (2005).

to longevity is constant as individuals get older, and is a linear function of years saved. The traditional age-invariant VSL value is, of course, constant as individuals get older and does not vary with years saved. The age-adjusted VSL value may not vary much as individuals get older, and does not vary with years saved.

The fact that age-adjusted VSLs do not vary much is initially surprising, but not on reflection. If age-adjusted VSLs were a perfect proxy for LWUs, then these values would decrease with age, as long as life expectancy is decreasing.²⁴¹ But cognitive errors, the dead-anyway effect, shorter investment horizons for the aged, and other distorting effects (relative to LWUs) all mean that age-adjusted VSLs need not decrease with age, even when life expectancy does. Alberini et al., in a contingent-valuation study, found that WTP to avoid mortality risk did not decline with age among Americans, even after age seventy.²⁴² Kerry Smith et al., in one estimation based on wage data, found that VSL *increases* with age.²⁴³ Aldy and Viscusi, in a more recent wage study, estimate an “inverted U” relationship, with VSL increasing and then decreasing with age. For example, individuals aged eighteen to twenty-two have a VSL of \$3.13 million; VSL *increases* until age twenty-eight to thirty-two (where it equals \$5.76 million) and then decreases, reaching \$2.51 million for individuals aged fifty-eight to sixty-two.²⁴⁴

Finally, if age-adjusted VSLYs are calculated by dividing the age-adjusted VSL by discounted life-expectancy, these amounts, too, will be an “inverted U” function of age,²⁴⁵ and, for a given age, will be a linear function of years saved.²⁴⁶

241. More precisely, the LWU value of life remaining should decrease with age, assuming life expectancy decreases and the value of background characteristics does not change.

242. Alberini et al., *supra* note 104, at 771. WTP declined among Canadians after age seventy.

243. V. Kerry Smith et al., *Do the Near-Elderly Value Mortality Risks Differently?*, 86 REV. ECON. & STAT. 423, 427-28 (2004).

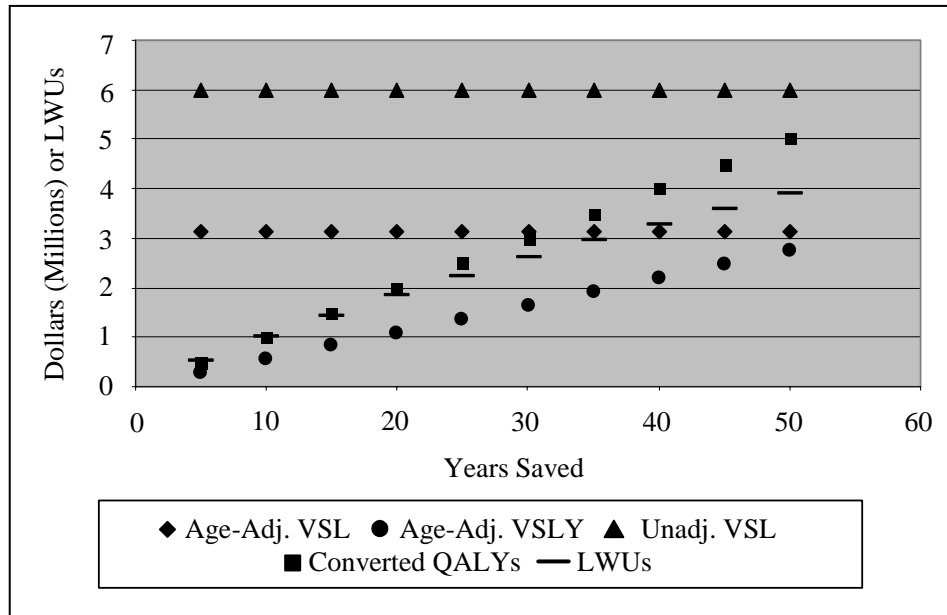
244. See Joseph E. Aldy & W. Kip Viscusi, *Age Variations in Workers' Value of Statistical Life* 19-23, 42, 49-50 (Nat'l Bureau of Econ. Research, Working Paper No. 10,199, 2003), available at <http://www.nber.org/papers/w10199>. For reviews of the literature on how VSL varies with age, see *id.* at 1-4; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 992-94; and Revesz & Stavins, *supra* note 191, at 21. Two recent studies are Thomas J. Kniesner et al., *Life-Cycle Consumption and the Age-Adjusted Value of Life* (Harvard John M. Olin Ctr. for Law, Econ., & Bus., Discussion Paper No. 459, 2004), available at <http://ssrn.com/abstract=580761>; and Anna Alberini et al., *Willingness To Pay To Reduce Mortality Risks: Evidence from a Three-Country Contingent Valuation Study* (Fondazione Eni Enrico Mattei, Working Paper No. 2004.111, 2004), available at <http://www.feem.it/NR/rdonlyres/8904A715-57A3-4FDD-A7E9-52318537EEFF/1258/11104.pdf>.

245. See Aldy & Viscusi, *supra* note 244, at 23-24.

246. The age-adjusted VSLY technique, in its simplest variant, determines an age-specific VSLY for each age, and then monetizes the change to someone's longevity induced by a policy

The following graphs suggest how LWUs, and the different proxies for LWUs, might correlate with longevity and age.²⁴⁷

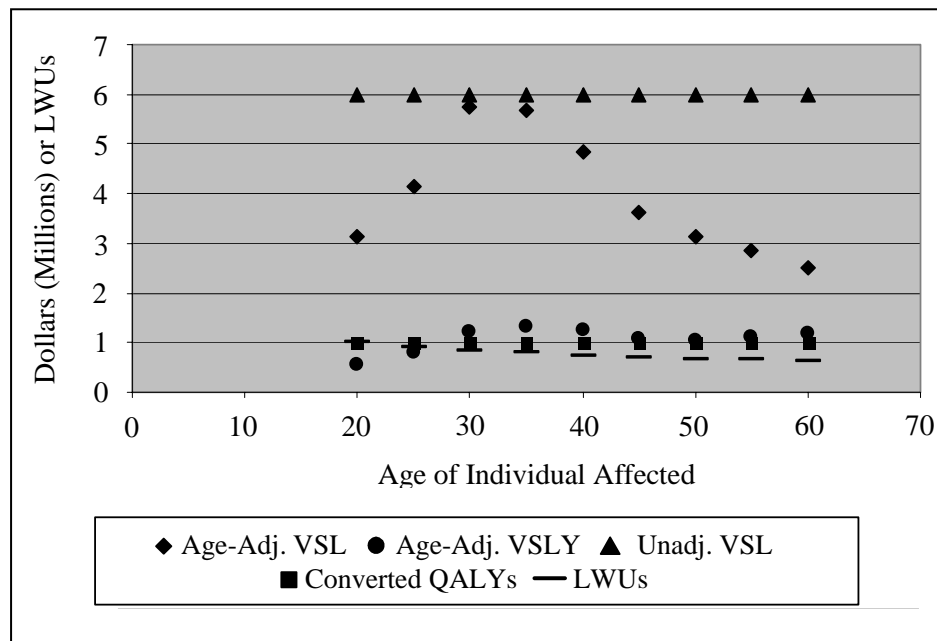
FIGURE 5: VALUATIONS OF INCREASING INCREMENTS TO LONGEVITY FOR INDIVIDUALS OF CONSTANT AGE (20 YEARS OLD)



choice by multiplying the life-years saved or lost by a single age-adjusted VSLY determined by the individual's age in the status quo. It is this variant of the approach that is linear in changes to longevity, and that I focus on in this Section. More complicated variants would incorporate a discount rate into the policy analysis or use a different age-specific VSLY for each year added or lost (so that the first incremental year of a policy that adds three years to the life of a fifty year-old would be valued at the age fifty VSLY, the second year at the age fifty-one VSLY, and the third year at the age fifty-two VSLY). I do not evaluate these refinements to the approach here.

247. In the first graph, the individuals are aged twenty, and five- to fifty-year increments to longevity are valued using converted QALYs (\$100,000 per QALY); unadjusted VSL (\$6 million); age-adjusted VSL (from Aldy & Viscusi, *supra* note 244, at 42); age-adjusted VSLYs (calculated by dividing the age-adjusted VSL by the age-specific life expectancy without discounting—since nondiscounted QALYs and LWUs are employed—and then multiplying by the increment to longevity); and LWUs (on the assumption that the LWU value of a life is the square root of its length). In the second graph, the individuals range from age twenty through sixty, and a ten-year increment to longevity is valued using the same measures.

FIGURE 6: VALUATIONS OF A CONSTANT (10 YEAR) INCREMENT TO LONGEVITY

TABLE 2: VARIANCE IN THE DOLLARS/LWU RATIO²⁴⁸

Varying Increments to Longevity (20 Years), Individual of Constant Age		Constant Increment to Longevity (10 Years), Individuals of Varying Age	
<i>Dollar Measure</i>	<i>Variance Log (Dollars/LWU) in Ascending Order</i>	<i>Dollar Measure</i>	<i>Variance Log (Dollars/LWU) in Ascending Order</i>
Converted QALY	0.0017	Converted QALY	0.0045
Age-Adj. VSLY	0.0017	Unadj. VSL	0.0045
Unadj. VSL	0.0683	Age-Adj. VSL	0.0124
Age-Adj. VSL	0.0683	Age-Adj. VSLY	0.0282

These analyses suggest that, for sets of choice situations where the age of persons whose longevity is affected by agency decisions is constant, but increments to longevity vary, QALYs and age-adjusted VSLYs will correlate more closely with LWUs than do age-invariant VSLs or age-adjusted VSLs. In other words, the variance in the QALY/LWU and VSLY/LWU ratio will be

248. For each set of longevity changes (constant age and increasing increments to longevity, or constant increments and increasing age), I calculated the log QALY/LWU, log VSL/LWU, log age-adjusted VSL/LWU, and log VSLY/LWU ratio for each change, and then calculated the variance.

lower than the variance in the VSL/LWU or adjusted VSL/LWU ratio.²⁴⁹ For sets of choice situations where increments to longevity are held constant but age varies, QALYs and age-invariant VSLs will correlate more closely with LWUs than do age-adjusted VSLYs and age-adjusted VSLs.

What if background characteristics are allowed to vary? Consider the simplest version of this case, where the individuals are healthy and vary in the value of their background characteristics but the longevity changes do not change the value of those characteristics. The LWU value for the loss or addition of a life-year to an individual will be the product of a person-specific scaling factor (for background characteristics) and an amount that decreases as the individual gets older.²⁵⁰ The QALY value for the loss or addition of a life-year to an individual will be constant, regardless of the individual's background characteristics and age. QALYs, in this context, are clearly imperfect proxies for LWUs. Age-invariant VSLs, age-adjusted VSLs, and age-adjusted VSLYs could clearly be better proxies for LWUs if they were adjusted for background characteristics: if individuals with greater consumption, or better sex lives, were given greater VSLs, age-adjusted VSLs, or age-adjusted VSLYs. Agencies do not thus adjust VSLs, and certainly should consider doing so.²⁵¹ Until they do, however, it would seem that hybrid cost-benefit analysis incorporating monetized QALYs is a better way to value lifesaving than the alternatives.

This conclusion, I should stress, is no more than an educated guess. Much more analysis remains to be done. My rough and ready treatment in this Section considered only the two limiting cases where the number of years saved or lost by agency action varies across choice situations but the age of those affected does not, or vice versa. More realistically, the instructions that OMB or other system-designers provide to agency decision-makers about the valuation of longevity will cover a more heterogeneous range of choice situations—where both the number of years saved or lost and the age of those affected, as well as their background characteristics, can vary. The most important point of this Section is not the substantive recommendation to use QALYs in valuing longevity, but is rather methodological: *None* of the alternatives on the table (QALYs, age-adjusted VSLYs, age-invariant VSLs, or age-adjusted VSLs) are perfect proxies for LWUs when it comes to changes in longevity, and the choice

249. This statement about variance assumes, of course, that QALYs have been converted to dollars using an appropriate conversion factor. Otherwise the QALY/LWU ratio could have a different variance from the VSLY/LWU or VSL/LWU variance simply because of the difference in units. In the variance tables above I have used the variance in the logarithm of the dollar/LWU ratio to wash out the scaling effect.

250. See *supra* text accompanying notes 100, 164-165.

251. See Cass R. Sunstein, *Valuing Life: A Plea for Disaggregation*, 54 DUKE L.J. 385, 386-89 (2004).

between them is a complicated exercise in optimization for OMB or other bodies (“system-designers”) that guide agency decision-makers in performing cost-benefit analysis.

E. Recommendations and Summary

The analysis in this Part has been quite complicated, so let me summarize the main prescriptions that emerge from it. Although, in principle, administrative decision-makers could seek to maximize welfare directly—by measuring the welfare effects of their choices on an LWU scale, rather than a dollar scale—this approach would amount to a radical change in current policy-analytic practice. I have therefore focused on a less radical possibility: nontraditional or “hybrid” cost-benefit analysis, where the scale for measuring policy impacts is a monetary scale, but the dollar amounts for certain welfare effects, such as health, are calculated by converting QALYs to dollars (or in some other nontraditional way) rather than by employing the traditional WTP/WTa methodology.

“Hybrid” cost-benefit analysis is no pipe dream. The FDA has used the approach in almost twenty rulemakings. This Part has described the FDA’s activities, and has taken a first stab at analyzing how hybrid cost-benefit analysis should be structured. To begin: At what rate should QALYs be converted to dollars? On this issue, I have criticized the FDA’s approach (which is to derive the conversion factor from a VSL value) and have argued that the choice of conversion factor is, rather, a pragmatic matter of maximizing expected LWUs. The system-designer who specifies the conversion factor (OMB, or the Administrator or policy office of a particular agency) will need to estimate, at least in a rough and ready way, how QALY measures of health effects and WTP/WTa measures of non-health effects correlate with LWUs, and should choose a QALY-to-dollar conversion factor that maximizes expected LWUs given these estimates. There is no natural rate at which QALYs convert to dollars. The conversion factor should be seen not as a mirror of some spurious economic reality, but rather as a numerical setting chosen to optimize the functioning of a decision-making tool, namely cost-benefit analysis.

What role should QALY-to-dollar conversions play in cost-benefit analysis? First, these numbers should be used in lieu of WTP/WTa amounts where WTP/WTa data are not available. Second, where both QALY and WTP/WTa measures of a particular type of health effect are available, agencies should conduct cost-benefit analysis using both measures. If the policy prescriptions resulting from the parallel analyses differ, the agency should reexamine its WTP/WTa data. (In effect, then, the QALY measure functions as a rough test of the value of acquiring further WTP/WTa information.) If the divergence persists, then the agency will need to choose between the WTP/WTa and converted QALY measure. That choice, like the initial choice of conversion factor, is a

pragmatic one—a matter of maximizing expected LWUs. System-designers should instruct agencies to be guided by hybrid over traditional cost-benefit analysis, or vice versa, depending on the designers’ estimates or rough guesses of the correlation between WTP/WTAs, QALYs, and LWUs.

For all the reasons discussed in Parts III and IV, neither QALYs nor WTP/WTAs are perfect proxies for LWUs. In some contexts, traditional cost-benefit analysis will be a superior tool for maximizing overall welfare. In other contexts—particularly where background characteristics such as wealth do not vary much among those affected by policy choice, or where available WTP/WTAs are not sensitive to background characteristics—nontraditional cost-benefit analysis will be superior. For reasons discussed in Section V.D., measuring the cost of death may well be a policy context where nontraditional cost-benefit analysis *is* superior. The existing literature on the choice between QALYs, age-invariant VSLs, age-adjusted VSLs, and age-adjusted VSLYs completely misses the crucial point that *none* of these measures are perfectly correlated with the welfare value of longevity. Longevity, like money and many other goods, quite likely has a diminishing marginal impact on individual and overall welfare. Thus the LWU value of a given life expectancy decreases as individuals age, but the QALY and age-invariant VSL values remain constant, and the age-adjusted VSL or VSLY values may increase.²⁵² And both age-invariant and age-adjusted VSLs, by contrast with QALYs and LWUs, are not sensitive to changes in the amount of life expectancy holding age constant. This admittedly preliminary analysis suggests that, absent adjustments to VSL or VSLY values for wealth or other background characteristics, QALYs are probably a better measure of the value of lifesaving than alternatives.

CONCLUSION

This Article has provided a novel, *welfarist* view of QALYs. Although the academic literature on QALYs is huge, encompassing a wealth of survey data, cost-effectiveness research, and ancillary analysis,²⁵³ QALYs have (until recently) been little used by governmental bodies in the United States²⁵⁴—in part because their policy role has been poorly understood. Welfare economists and other welfarists, who think that well-being-related constructs such as Kaldor-

252. As mentioned earlier, age-adjusted VSLs and age-adjusted VSLYs both may have an “inverted U” shape, first increasing and only later decreasing with age. *See supra* text accompanying notes 244-246. If, for example, the age-adjusted VSL and VSLY for a fifty year-old are greater than for a forty year-old, both methods will place a larger money value on an X-year increment to the fifty year-old’s expected longevity than on the same, X-year increment to the forty year-old’s expected longevity.

253. *See supra* text accompanying notes 3-7.

254. *See supra* text accompanying notes 8-20.

Hicks efficiency or overall welfare should play a large role in determining governmental choices, will be unpersuaded by the “extrawelfarist” perspective on QALYs that dominates the public health literature. Welfarists wonder: Why should a policymaker ever use QALYs rather than a WTP/WTa scale to measure health benefits?²⁵⁵ I have offered an innovative answer to that question, making claims along the way that will undoubtedly surprise many welfarists.

In particular (here building on my prior work with Eric Posner), I have argued that policymakers should focus on overall well-being, not Kaldor-Hicks efficiency, and that cost-benefit analysis is no normative gold standard, but simply a decision-making tool that can help policymakers maximize overall welfare.²⁵⁶ I have suggested that overall well-being is, in principle, measurable in LWUs (lifetime welfare units), and that QALYs and WTP/WTa amounts are both imperfect, practicable estimates of LWUs.²⁵⁷ WTP/WTa valuations deviate from LWUs for a variety of reasons—wealth effects, the “dead anyway” effect, risk and tradeoff biases, and others—and although QALYs certainly have their own flaws they can, in some contexts, furnish an improved scale of welfare. I have proposed that QALYs should function, not as the nonmonetary maximand in a cost-effectiveness analysis, but rather as a valuation of health that is converted into dollars using some conversion factor and then incorporated into a monetized cost-benefit analysis. And I have analyzed how this nontraditional or “hybrid” cost-benefit analysis should be structured: in particular, what the optimal QALY-to-dollar conversion rate is, and when QALY-to-dollar conversions should displace WTP/WTa amounts as the monetary measure of health.²⁵⁸

255. *See supra* text accompanying notes 31-59 (discussing welfarist and extrawelfarist views of QALYs).

256. *See supra* text accompanying notes 57-59.

257. *See supra* Parts II-IV.

258. *See supra* Part V.

Appendix I: Cost-Effectiveness Analysis, and How It Can Deviate from Traditional Cost-Benefit Analysis²⁵⁹

The following example illustrates the two variants of cost-effectiveness analysis (fixed budget and cut-off ratio), using QALYs as the metric of effectiveness, and demonstrating how both variants can deviate from traditional cost-benefit analysis.

Imagine that a governmental body can implement two general programs, which are not mutually exclusive. Each program can be implemented through a variety of mutually exclusive subprograms; if one subprogram within the group is picked, another cannot be.²⁶⁰

A subprogram is “dominated” by another within the same general program if the first has smaller total effectiveness (QALYs) and greater total costs.²⁶¹ A dominated subprogram can be eliminated from consideration; nothing is lost, along either the health or cost dimension, by replacing a dominated subprogram with the one that dominates it.

Table 1 shows the total dollar cost of each subprogram as well as its total QALY benefit, with “dominated” subprograms eliminated. The subprograms are listed in order of total effectiveness. To be clear, the numbers in this table are the total costs for each possible subprogram, relative to the status quo of inaction. For example, if subprogram *A* is picked, the total cost will be \$1 million and the total QALY benefit will be 10 QALYs. If subprogram *B* is picked, the total cost will be \$2 million and the total QALY benefit will be 14 QALYs. *A* and *B* are mutually exclusive, and thus cannot be jointly picked. However, *A* and *F* are not mutually exclusive. If they are jointly picked, the total cost of that policy, relative to the status quo, will be \$3 million and the total QALY benefit will be 22 QALYs.

259. The numerical example used in this section is based on Karlsson & Johannesson, *supra* note 26.

260. By contrast with standard presentations of cost-effectiveness analysis, my example assumes, for simplicity, that the subprograms are indivisible—they cannot be partially implemented. For each program, the decision-maker has the choice of not implementing any of its subprograms or fully implementing only one of its subprograms.

261. Strictly, this should be: Subprogram *PI* is dominated by subprogram *P2* if (1) *PI*’s health benefits are less than *P2*’s and *PI*’s costs are greater than or equal to *P2*’s, or (2) *PI*’s health benefits are less than or equal to *P2*’s and *PI*’s costs are greater than *P2*’s.

TABLE 1: TOTAL COSTS AND QALYs OF POSSIBLE SUBPROGRAMS

<i>Program I</i>			<i>Program II</i>		
Subprograms	Total Cost (millions)	Total Benefit (QALYs)	Subprograms	Total Cost (millions)	Total Benefit (QALYs)
<i>A</i>	\$1	10	<i>F</i>	\$2	12
<i>B</i>	\$2	14	<i>G</i>	\$4	16
<i>C</i>	\$3	16	<i>H</i>	\$5.5	18
<i>D</i>	\$4	19			
<i>E</i>	\$5	20			

The next table shows the incremental cost, effectiveness, and cost-effectiveness ratios of each subprogram, relative to the subprogram above it in the table.

TABLE 2: INCREMENTAL COST-EFFECTIVENESS OF SUBPROGRAMS

<i>Program I</i>				<i>Program II</i>			
Sub-programs	Δ Cost (millions)	Δ QALYs	Δ Cost/ Δ QALYs	Sub-programs	Δ Cost (millions)	Δ QALYs	Δ Cost/ Δ QALYs
<i>A</i>	\$1	10	\$100,000	<i>F</i>	\$2	12	\$166,667
<i>B</i>	\$1	4	\$250,000	<i>G</i>	\$2	4	\$500,000
<i>C</i>	\$1	2	\$500,000	<i>H</i>	\$1.5	2	\$750,000
<i>D</i>	\$1	3	\$333,333				
<i>E</i>	\$1	1	\$1,000,000				

The fixed-budget variant of cost-effectiveness analysis tells the decision-maker to choose that mix of subprograms which maximizes QALYs for the given budget. Assume that, in the case at hand, the budget to be maximized is \$4 million. Then the QALY-maximizing mix of subprograms is *B* and *F*. No other mix of subprograms that costs less than or equal to \$4 million produces more QALYs.²⁶²

What about the cut-off ratio variant of cost-effectiveness analysis? The decision rule here is as follows. “Weakly dominated” subprograms are eliminated from consideration. A subprogram is “weakly dominated” if its incremental cost-effectiveness ratio is greater than the incremental cost-effectiveness ratio of the subprogram immediately below it in the table. In the current example, subprogram *C* is weakly dominated. Then, incremental cost-effectiveness ratios for the remaining subprograms are recalculated, as shown in Table 3. Finally,

262. In the simplified case at hand, the budget-maximizing mix of subprograms can be identified by inspection. In more complicated cases, if certain assumptions are made about the divisibility of subprograms and the constancy of returns to scale, the budget-maximizing mix can be identified using a decision rule that looks to incremental cost-effectiveness ratios. See Johansson, *supra* note 26, at 484-85.

“the [subprogram] within each cluster that has the highest incremental cost-effectiveness ratio that is equal to or below the [cut-off ratio] should be implemented.”²⁶³ For example, if the cut-off ratio is \$200,000 per QALY, then the government applying this decision rule picks subprograms *A* and *F*.

TABLE 3: INCREMENTAL COST-EFFECTIVENESS OF SUBPROGRAMS, EXCLUDING WEAKLY DOMINATED (SUBPROGRAM C)

<i>Program I</i>				<i>Program II</i>			
Sub-programs	Δ Cost (millions)	ΔQALYs	ΔCost/ΔQALYs	Sub-programs	Δ Cost (millions)	ΔQALYs	ΔCost/ΔQALYs
<i>A</i>	\$1	10	\$100,000	<i>F</i>	\$2	12	\$166,667
<i>B</i>	\$1	4	\$250,000	<i>G</i>	\$2	4	\$500,000
				<i>H</i>	\$1.5	2	\$750,000
<i>D</i>	\$2	5	\$400,000				
<i>E</i>	\$1	1	\$1 million				

To see how both variants of cost-effectiveness analysis can deviate from traditional cost-benefit analysis, consider the following table, which shows the total costs and the total QALYs as well as total WTP for the different subprograms in Table 1—on the assumption that Program I benefits a population all members of which are willing to pay a constant \$300,000 per QALY, while Program II benefits a population all members of which are willing to pay only a constant \$100,000 per QALY.

TABLE 4: PROGRAM COSTS AND PROGRAM BENEFITS (BENEFICIARIES OF PROGRAM I ARE WILLING TO PAY \$300,000 PER QALY, WHILE BENEFICIARIES OF PROGRAM II ARE WILLING TO PAY \$100,000 PER QALY)

<i>Program I</i>				<i>Program II</i>			
Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)	Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)
<i>A</i>	\$1	10	\$3	<i>F</i>	\$2	12	\$1.2
<i>B</i>	\$2	14	\$4.2	<i>G</i>	\$4	16	\$1.6
<i>C</i>	\$3	16	\$4.8	<i>H</i>	\$5.5	18	\$1.8
<i>D</i>	\$4	19	\$5.7				
<i>E</i>	\$5	20	\$6.0				

The fixed-budget variant of cost-effectiveness analysis, with a budget of \$4 million, tells the government to pick subprograms *B* and *F*. But aggregate WTP for the health benefits would be greater if the government spent the \$4 million on

263. Johannesson, *supra* note 26, at 484.

subprogram *D* instead (\$5.7 million versus \$5.4 million). The cut-off ratio variant, with a ratio of \$200,000 per QALY (the average of the two populations) tells the government to pick subprograms *A* and *F*—spending a total of \$3 million. But aggregate WTP would be increased by spending the \$3 million on subprogram *C* (\$4.8 million versus \$4.2 million). The best choice of all, as per traditional cost-benefit analysis, would be to spend only \$2 million on subprogram *B*. All of the Program II subprograms have greater total money costs than benefits. In Program I, by moving from subprogram *D* to *C* we save \$1 million in costs and give up only \$0.9 million in benefits; and by moving again from *C* to *B*, we save \$1 million in costs and give up only \$0.6 million in benefits.

Assume now that all individuals in both populations are willing to pay a constant \$200,000 per QALY. In that case, the costs and benefits of the various subprograms are as follows:

TABLE 5: PROGRAM COSTS AND PROGRAM BENEFITS (ALL BENEFICIARIES ARE WILLING TO PAY \$200,000 PER QALY)

<i>Program I</i>				<i>Program II</i>			
Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)	Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)
A	\$1	10	\$2	F	\$2	12	\$2.4
B	\$2	14	\$2.8	G	\$4	16	\$3.2
C	\$3	16	\$3.2	H	\$5.5	18	\$3.6
D	\$4	19	\$3.8				
E	\$5	20	\$4.0				

As before, the fixed-budget variant of cost-effectiveness analysis maximizes QALYs for a budget of \$4 million by picking subprograms *B* and *F*. As before, the cut-off ratio variant using the \$200,000 per QALY cut-off selects subprograms *A* and *F*. But this time, with WTP constant at \$200,000 per QALY, traditional cost-benefit analysis also selects *A* and *F*, as can be seen by inspection.

Appendix II: Hybrid Cost-Benefit Analysis²⁶⁴

I. THE OPTIMAL QALY-TO-DOLLAR CONVERSION FACTOR

When choosing between the status quo and a project that will yield q QALYs in health benefits while costing w dollars,²⁶⁵ a decision-maker given a QALY-to-dollar conversion factor f will implement the project if and only if $fq > w$. The system-designer, having a subjective probability distribution with respect to the levels of health benefits in QALYs and non-health costs in WTPs, as well as their respective ratios to LWUs, should try to determine a fixed value of f that, when used by the decision-maker, will yield the greatest overall increase in LWUs.

A. The General Case

Let A_f be defined for all $f \in R^+$ as the set of pairs $\langle q, w \rangle \in (R^+)^2$, such that $fq > w$. In a choice situation where a project will cost w dollars and yield q QALYs in benefits, a decision-maker using a fixed QALY-to-dollar conversion factor, f , will select the project if and only if $\langle q, w \rangle \in A_f$.

Let $p(r, s, q, w)$ be the probability density function of continuous random variables R , S , Q , and W where

R represents the QALY/LWU ratio for health benefits,

S represents the WTP/LWU ratio for non-health costs,

Q represents the health benefits, measured in QALYs, and

W represents the non-health costs, measured in WTP.

Then, the expected change in LWUs from the implementation of the project using any conversion factor, f , is

$$\iiint_{A_f} \iiint_{R^2} \left(\frac{q}{r} - \frac{w}{s} \right) p(r, s, q, w) dr ds dq dw;$$

and so the optimal QALY-to-dollar conversion factor will be the value of f that

264. Many thanks to Craig Phillips for preparing this Appendix.

265. Assume here that all of the costs are non-health costs, directly measurable in WTPs, and that all of the benefits are health benefits, directly measurable in QALYs.

maximizes the above function.

B. If r and s are independent of q and w , and if r is independent of s , then the optimal value of f will be $f_0 = \left(\frac{E(\frac{1}{r})}{E(\frac{1}{s})} \right)$.

Because of the above independence assumptions, p can be represented as the product of p_1 , the density function of R , p_2 , the density function of S , and p_3 , the density function of Q and W .

$$\begin{aligned} & \iint_{A_f} \iint_{R^2} \left(\left(\frac{q}{r} \right) - \left(\frac{w}{s} \right) \right) p_1(r) p_2(s) p_3(q, w) dr ds dq dw \\ &= \iint_{A_f} \left(q \int_R \frac{1}{r} p_1(r) dr - w \int_S \frac{1}{s} p_2(s) ds \right) p_3(q, w) dq dw \\ &= \iint_{A_f} \left(q E\left(\frac{1}{r}\right) - w E\left(\frac{1}{s}\right) \right) p_3(q, w) dq dw. \end{aligned}$$

When the conversion factor is set at f_0 , the increase in overall LWUs is determined by integrating over the set A_{f_0} , which is the set of all $\langle q, w \rangle \in (R^+)^2$ such that $q \left(\frac{E(\frac{1}{r})}{E(\frac{1}{s})} \right) > w$ or, equivalently, where $q E\left(\frac{1}{r}\right) - w E\left(\frac{1}{s}\right) > 0$. The expected increase in overall LWUs thus reaches its maximum when the decision-maker uses f_0 as the conversion factor.

C. When (1) the probability density function, u , of the log of the QALY/LWU ratio is symmetric about its mean, $\log r_0$, increasing only when $r < r_0$; (2) the WTP/LWU ratio for costs is fixed; and (3) the value of the probability density function for costs and benefits, $v(c, b)$, is such that $v(x, y) = v(y, x)$, the optimal conversion factor, f_ , is $\frac{s}{r_0}$, where s is the fixed WTP/LWU ratio for costs.*

Let

- s be the fixed WTP/LWU ratio for costs,
- r_0 be such that u has its mean at $\log r_0$,
- f be the QALY-to-dollar conversion factor,
- R be the continuous random variable representing the QALY/LWU ratio,
- $u(\log r)$ be the probability density function of $\log R$,
- u be independent of v and increasing when $r < r_0$ and decreasing otherwise,
- $U(x)$ be the cumulative distribution function for $\log R$, and
- the probability density function of the costs and benefits in LWUs, $v(c, b)$, be such that, for all $x, y \in R^+$, $v(x, y) = v(y, x)$.

Let $x, y \in R^+$, such that $x \geq y$.

When $b = x$ and $c = y$, there will be a loss of $x - y$ whenever $frx < sy$, or $r < \frac{sy}{fx}$.

When $b = y$ and $c = x$, there will be a loss of $x - y$ whenever $fry > sx$, or $r > \frac{sx}{fy}$.

If there is a single QALY-to-dollar conversion factor that minimizes the sum of the probabilities that $r < \frac{sy}{fx}$ and that $r > \frac{sx}{fy}$ for all $x, y \in R^+$, then that factor will be the optimal conversion factor.

The probability that $r < \frac{sy}{fx}$ is

$$\int_{-\infty}^{\log \frac{sy}{fx}} u(\log r) d(\log r) = \int_{-\infty}^1 u(\log r) d(\log r) + \int_1^{\log \frac{sy}{fx}} u(\log r) d(\log r),$$

and, similarly, the probability that $r > \frac{sx}{fy}$ is

$$\int_{\log \frac{sx}{fy}}^{+\infty} u(\log r) d(\log r) = \int_{\log \frac{sx}{fy}}^1 u(\log r) d(\log r) + \int_1^{+\infty} u(\log r) d(\log r).$$

In order to find the value of f that minimizes these probabilities, we differentiate the sum of these two probabilities with respect to f , so that

$$\begin{aligned} & \frac{d}{df} \left(\int_{\log \frac{sx}{fy}}^1 u(\log r) d(\log r) + \int_1^{\log \frac{sy}{fx}} u(\log r) d(\log r) \right) \\ &= \frac{d}{df} \left(\int_{\log \frac{sx}{fy}}^{\log \frac{sy}{fx}} u(\log r) d(\log r) \right) \\ &= \frac{d}{df} \left(U\left(\log \frac{sy}{fx}\right) - U\left(\log \frac{sx}{fy}\right) \right) = -\frac{1}{f} \left(u\left(\log \frac{sy}{fx}\right) - u\left(\log \frac{sx}{fy}\right) \right). \end{aligned}$$

Let f_* be such that $u\left(\log \frac{sy}{f_*x}\right) = u\left(\log \frac{sx}{f_*y}\right)$. It follows from the above that the sum of (1) the probability of error when costs are x and benefits are y and (2) the probability of error when costs are y and benefits are x reaches its minimum for all $\langle x, y \rangle \in (R^+)^2$ when $f = f_*$. Because $u(\log r)$ is symmetrical about its mean, $\log r_0$, it follows that if $f_* = \frac{s}{r_0}$, then:

$$u\left(\log \frac{sy}{f_*x}\right) = u\left(\log \frac{r_0y}{x}\right) = u\left(\log(r_0) + \log\left(\frac{y}{x}\right)\right) = u\left(\log(r_0) - \log\left(\frac{y}{x}\right)\right) = u\left(\log \frac{sx}{f_*y}\right).$$

The optimal WTP/QALY conversion factor is therefore $f_* = \frac{s}{r_0}$.

II. CHOOSING BETWEEN QALYS AND WTPs AS THE MEASURE OF HEALTH BENEFITS

Let the WTP/LWU ratio for costs of a set of projects be fixed at s and let the random variables R_w and R_q represent the WTP/LWU and converted QALY/LWU ratios for the health benefits, respectively. Assume that R_w and R_q are independent of the levels of costs and benefits in LWUs. Then, if the probability density functions, $p_w(r)$ of R_w and $p_q(r)$ of R_q , are such that $\int_{-\infty}^x p_q(r)dr < \int_{-\infty}^x p_w(r)dr$ for every $x < s$ and $\int_x^{+\infty} p_q(r)dr < \int_x^{+\infty} p_w(r)dr$ for every $x > s$, then using a converted QALY measurement of benefits, rather than a direct dollar measurement, will yield greater expected overall welfare in LWUs.

If $c > b$, then the project will be implemented at a loss if $sc < rb$, or $r > \frac{sc}{b}$, when r is the ratio of the monetized benefit measure (WTP or converted QALYs) to LWUs. By hypothesis, because $\frac{sc}{b} > s$, the probability that the project will be implemented at a loss is greater when measuring benefits directly in WTP than with converted QALYs. Analogous reasoning yields the same result where $c < b$, and so measuring benefits in converted QALYs will yield greater expected overall LWUs than a direct dollar measurement.

Conversely, if $\int_{-\infty}^x p_w(r)dr < \int_{-\infty}^x p_q(r)dr$ for every $x < s$ and

$\int_x^{+\infty} p_w(r)dr < \int_x^{+\infty} p_q(r)dr$ for every $x > s$, then, by analogous reasoning,

using a direct dollar measurement of benefits will yield greater overall welfare in LWUs than a converted QALY measurement will yield.